Exp Mail EV449562444US
USAN 10/648,854

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07K 13/00, 15/28, C07H 21/04 C12N 5/10, C12P 21/00 A61K 37/02, 39/395, 48/00 G01N 33/53	A1	(11) International Publication Number: WO 94/11401 (43) International Publication Date: 26 May 1994 (26.05.94)

(21) International Application Number: PCT/US93/11097

(22) International Filing Date: 16 November 1993 (16.11.93)

(30) Priority data:
978,897 17 November 1992 (17.11.92) US

(71) Applicant: YALE UNIVERSITY [US/US]; 451 College Street, New Haven, CT 06520 (US).

(72) Inventors: RIMM, David, L. ; 7 Weir Street, Branford, CT 06405 (US). MORROW, Jon, S. ; 85 Soundview Avenue, Madison, CT 06443 (US).

(74) Agents: MISROCK, S., Leslie et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AU, BB, BG, BR, BY, CA, CZ, FI, HU, JP, KR, KZ, LK, LV, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: HUMAN HOMOLOG OF THE E-CADHERIN GENE AND METHODS BASED THEREON

(57) Abstract

The present invention relates to nucleotide sequences of the human E-cadherin gene, and the amino acid sequences of the encoded E-cadherin protein. The invention further relates to fragments and other derivatives, and analogs, of the human E-cadherin protein, as well as antibodies thereto. Nucleic acids encoding such fragments or derivatives are also within the scope of the invention, as well as human E-cadherin antisense nucleic acids. Production of the foregoing proteins and derivatives, e.g., by recombinant methods, is provided. In specific embodiments, the invention relates to human E-cadherin protein derivatives and analogs of the invention which are functionally active, or which comprise one or more domains of a human E-cadherin protein, including but not limited to the amino-terminal processed region, the HAV homotypic binding domain, one or more of the three repeat domains, the transmembrane region, the extracellular region, the cytoplasmic domain, and any combination of the foregoing. The present invention further relates to therapeutic and diagnostic methods and compositions based on E-cadherin proteins and nucleic acids.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovakia
SN	Senegal
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

HUMAN HOMOLOG OF THE E-CADHERIN GENE AND METHODS BASED THEREON

This invention was made with government support under grant number 1RO1 DK43812-01 awarded by the National Institutes of Health. The government has certain rights in the invention.

1. INTRODUCTION

The present invention relates to the human epithelial-cadherin (E-cadherin) gene and its encoded protein product, as well as derivatives and analogs of the human E-cadherin protein. Production of human E-cadherin proteins, derivatives and antibodies is also provided. The invention further relates to therapeutic and diagnostic methods and compositions.

2. BACKGROUND OF THE INVENTION

Cell adhesion molecules (CAMs) are cell surface glycoproteins which mediate specific cell-cell adhesions involved in embryonic development and maintaining tissue form and function.

E-cadherin is a cell adhesion molecule that is also known as uvomorulin, L-CAM and Cell CAM 120/80. E-cadherin localizes to the lateral surfaces and is concentrated in the adherens junctions of intestinal epithelial cells. It is present in epithelial cells from all organs examined, and related "cadherin family" molecules have been identified in brain, muscle, placenta, and other organs. This molecule has been attributed to play a role in initiation of the formation of the cortical cytoskeleton, establishment of polarity (Nelson, 1991, Sem. Cell. Bio. 2:375-385; Nelson and Hammerton, 1989, J. Cell Biol. 108:893-902; Nelson et al., 1990, J. Cell Biol. 110:349-357; McNeill et al., 1990, Cell 62:309-316; Pasdar et al., 1991, J. Cell Biol. 113:645-655; Ruggieri et al., 1992, Am. J. Pathol. 140:1179-1185; Avner et al., 1992, Proc. Natl. Acad. Sci. USA 89:7447-7451), and suppression of cell invasion (Behrens et al., 1989, J. Cell Biol. 108:2435-2447). It has recently gained increased attention since it has been proposed to be a tumor suppressor protein (Mareel et al., 1991,

Int. J. Cancer 47:922-928). Although this claim has yet to be proven in actual human tumors, there is evidence of changes in levels and patterns of expression in breast, gastric, and esophageal squamous cell carcinomas (Shimoyama and Hirohashi, 1991, Cancer Res. 51:2185-2192; Shimoyama et al., 1989, Cancer Res. 49:2128-2133; Shiozaki et al., 1991, Am. J. Pathol. 139:17-23).

Studies, primarily on the mouse and chicken proteins, have shown that E-cadherin is a 120 kilodalton (kD) membrane spanning protein with a large, glycosylated, amino-terminal extracellular domain and a 150 amino acid C-terminal cytoplasmic tail. The extracellular domain is cleavable with trypsin in the presence of Ca^{++}, resulting in an 80 kD peptide that contains three putative repeat structures (possibly involved in Ca^{++} binding) and a highly conserved amino-terminal 113 amino acid region. Within this region is a HAV (His-Ala-Val) motif that is conserved between species, and that is primarily responsible for homotypic interactions (Fig. 1). In mouse and chicken, an amino-terminal piece consisting of 150 amino acids is removed in a maturation process resulting in the truncated mature form. The exact mechanism of this processing event is unknown.

The murine and avian homologs of E-cadherin have been cloned and sequenced (see, *e.g.*, Ringwald et al., 1987, EMBO J. 6:3647-3653; Nagafuchi et al., 1987, Nature 329:341-343; Gallin et al., 1987, Proc. Natl. Acad. Sci. USA 84:2808-2812; Sorkin et al., 1988, Proc. Natl. Acad. Sci. USA 85:7617-7621). A partial sequence of human E-cadherin derived from a liver cDNA clone has been reported (Mansouri et al., 1988, Differentiation 38:67-71). A partial sequence derived from amino-terminal sequencing of the human protein has also been disclosed (Wheelock et al., 1987, J. Cell Biochem. 34:187-202). However, the full-length nucleotide and amino acid sequences for human E-cadherin have not been available prior to the present invention. A knowledge of such sequences is of primary importance since proteins having human sequences and antibodies thereto are greatly preferred over those of other species for human therapeutic and diagnostic purposes. Knowledge of the complete E-cadherin sequences is also important for deriving appropriate strategies in the

generation of derivatives and fragments, for example, in choosing an appropriate restriction enzyme for cleavage in the isolation of portions of the coding sequence.

Citation of a reference hereinabove shall not be construed as an admission that such reference is prior art to the present invention.

3. SUMMARY OF THE INVENTION

The present invention relates to nucleotide sequences of the human E-cadherin gene, and the amino acid sequences of the encoded E-cadherin protein. The invention further relates to fragments and other derivatives, and analogs, of the human E-cadherin protein, as well as antibodies thereto. Nucleic acids encoding such fragments or derivatives are also within the scope of the invention. Production of the foregoing proteins and derivatives, *e.g.*, by recombinant methods, is provided.

In specific embodiments, the invention relates to human E-cadherin protein derivatives and analogs of the invention which are functionally active, or which comprise one or more domains of a human E-cadherin protein, including but not limited to the amino-terminal processed region, the HAV homotypic binding domain, one or more of the three repeat domains, the conserved cysteine domain, the transmembrane region, the extracellular region, the cytoplasmic domain, and any combination of the foregoing.

The present invention further relates to therapeutic and diagnostic methods and compositions based on E-cadherin proteins and nucleic acids. The invention provides for treatment of disorders of cell fate or differentiation by administration of a therapeutic compound of the invention. Such therapeutic compounds (termed herein "Therapeutics") include: E-cadherin proteins and analogs and derivatives (including fragments) thereof; antibodies thereto; nucleic acids encoding the E-cadherin proteins, analogs, or derivatives; and E-cadherin antisense nucleic acids. In a preferred embodiment, a Therapeutic of the invention is administered to treat or prevent a cancerous condition, or to prevent progression from a pre-neoplastic or non-malignant state into a neoplastic or a

malignant state, or to inhibit or ameliorate metastatic tumor development. Methods of promoting nerve or tissue regeneration, of promoting wound healing, of treating an inflammatory disorder, and of treating or preventing gestational disease or fetal wastage are also provided.

In particular embodiments presented by way of examples sections *infra*, the invention provides the complete nucleotide sequence of human cDNA from liver and colon coding for E-cadherin, and the sequence of the encoded human E-cadherin protein. Also described are specific anti-human E-cadherin antibodies, and multiple human cloned fragments of both liver and colon E-cadherin cDNAs, some of which have been expressed in both eukaryotic and prokaryotic expression systems.

4. DESCRIPTION OF THE FIGURES

Figure 1. A schematic diagram of E-cadherin showing structural features including the amino (N)-terminal processed region, the HAV homotypic binding sequence, the three repeat domains, and the highly conserved carboxy (C)-terminal cytoplasmic domain. Also shown are some of the proteins thought to interact, either directly or indirectly, with the cytoplasmic domain.

Figure 2. A restriction map of the liver E-cadherin clone. The translation start is shown preceding the mature N terminus of the protein. Three proposed repeat domains are shown and the homotypic adhesion sequence (HAV) is located in the first repeat. The hashed portion of the sequence denotes the region originally published by Mansouri et al. (1988, Differentiation 38:67-71). The regions labeled e250 and cyto 20 are regions that have been produced as fusion proteins for in vitro binding studies and antibody production.

Figure 3. The complete nucleotide (SEQ ID NO:1) and protein (SEQ ID NO:2) sequences of human liver E-cadherin.

Figure 4. Series of E-cadherin clones shown by location and source. Leftmost column lists the clone name. Numbers beneath lines are the starting and ending nucleotide of each clone. Clones with hashmarks on one end are either clones whose ends are not yet sequenced or fusion clones with

artefactual sequence on the end as shown. Part A: liver clones. Part B: colon clones. Solid bar: characterized E-cadherin sequence; cross-hatched area: concatenated unrelated sequence; dotted area: regions not yet sequenced.

Figure 5. Nucleotide (SEQ ID NO:1) and protein (SEQ ID NO:2) sequences of human liver E-cadherin, with restriction sites. Restriction enzyme cleavage sites, translation start site, mature amino (N) terminus, homotypic binding domain/recognition sequence ("Rec Seq"; containing the HAV site), and the transmembrane region are shown.

Figure 6. Schematic diagram of plasmid pCMV-NeoPoly 1. pCMV-NeoPoly 1 is a 6.7 kb plasmid that was constructed and kindly provided by Dr. Eric R. Fearon. Known unique restriction sites: XhoI, EcoRV, BamHI, StuI, NheI, HindIII, and SstI.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to nucleotide sequences of the human E-cadherin gene, and the amino acid sequence of the encoded E-cadherin protein. The invention further relates to fragments and other derivatives, and analogs, of the human E-cadherin protein. Nucleic acids encoding such fragments or derivatives are also within the scope of the invention. Production of the foregoing proteins and derivatives, *e.g.*, by recombinant methods, is provided.

The invention also relates to human E-cadherin protein derivatives and analogs of the invention which are functionally active, *i.e.*, they are capable of displaying one or more known functional activities associated with a full-length (wild-type) E-cadherin protein. Such functional activities include but are not limited to antigenicity [ability to bind (or compete with a E-cadherin protein for binding) to an anti-E-cadherin protein antibody], immunogenicity (ability to generate antibody which binds to a E-cadherin protein), ability to bind (or compete with a E-cadherin protein for binding) to a receptor or ligand for a E-cadherin protein, suppression of cell invasiveness, therapeutic activity, etc.

The invention further relates to fragments (and derivatives and analogs thereof) of a human E-cadherin protein which comprise one or more domains of a human E-cadherin protein (see Section 6), including but not limited to the amino-terminal processed region, the HAV homotypic binding domain, one or more of the three repeat regions, the conserved cysteine domain, the extracellular region, transmembrane region, cytoplasmic domain, and any combination of the foregoing.

Antibodies to the human E-cadherin protein and its derivatives and analogs are additionally provided.

The present invention further relates to therapeutic and diagnostic methods and compositions based on E-cadherin proteins and nucleic acids. The invention provides for treatment by administration of a therapeutic compound of the invention. Such therapeutic compounds (termed herein "Therapeutics") include: human E-cadherin proteins and analogs and derivatives (including fragments) thereof; antibodies thereto; nucleic acids encoding the human E-cadherin proteins, analogs, or derivatives; E-cadherin antisense nucleic acids. In a preferred embodiment, a Therapeutic of the invention is administered to treat a cancerous condition, or to prevent progression from a pre-neoplastic or non-malignant state (*e.g.*, metaplastic condition) into a neoplastic or a malignant state, or to inhibit or ameliorate metastatic tumor development. In another specific embodiment, a nucleic acid encoding a human E-cadherin protein or fragment thereof is used in gene therapy. Methods of promoting nerve or tissue regeneration, of promoting wound healing, of treating an inflammatory disorder, and of treating or preventing gestational disease or fetal wastage are also provided.

E-cadherin plays a role in developmental and other physiological processes. The nucleic acid and amino acid sequences and antibodies thereto of the invention can also be used for the detection and quantitation of human E-cadherin mRNA, to study expression thereof, to produce human E-cadherin proteins, fragments and other derivatives, and analogs thereof, in the study and manipulation of differentiation and other physiological processes.

The invention is illustrated by way of examples *infra* which disclose, *inter alia*, the cloning and sequencing of human E-cadherin cDNAs from liver and colon, and the construction and recombinant expression of human E-cadherin chimeric/fusion derivatives and production of antibodies thereto.

For clarity of disclosure, and not by way of limitation, the detailed description of the invention will be divided into the following subsections:

(i) Isolation of the Human E-Cadherin Gene;
(ii) Expression of the Human E-Cadherin Gene;
(iii) Identification and Purification of the Expressed Gene Products;
(iv) Structure of the Human E-Cadherin Gene and Protein;
(v) Generation of Antibodies to the Human E-Cadherin Protein and Derivatives Thereof;
(vi) Human E-Cadherin Protein Derivatives and Analogs;
(vii) Assays of Human E-Cadherin Proteins, Derivatives and Analogs;
(viii) Therapeutic and Prophylactic Uses;
(ix) Gene Therapy;
(x) Antisense Regulation of Human E-cadherin Expression;
(xi) Demonstration of Therapeutic or Prophylactic Utility;
(xii) Therapeutic/Prophylactic Administration and Compositions;
(xiii) Diagnostic Utility.

5.1. ISOLATION OF THE HUMAN E-CADHERIN GENE

The invention relates to the nucleotide sequences of human E-cadherin nucleic acids. In a specific embodiment, the human E-cadherin nucleic acid comprises the nucleotide sequence (SEQ ID NO:1) shown in Figure 3; in particular, from nucleotides numbers 116-2749, or 566-2749, or 1037-2748, or fragments thereof. In other embodiments, a nucleic acid is provided which comprises the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1) from nucleotide numbers 1-1053, 510-2686, 1332-3000, 540-1500,

348-906, 890-1648, 384-1208, 641-2046, 685-1336, 880-1661, 1199-1742, 1373-1742, 1705-2204, or 2458-2775 (see Fig. 4).

In another specific embodiment, the nucleotide sequence encodes all or a portion of the amino acid sequence (SEQ ID NO:2) shown in Figure 3. The invention provides nucleic acids consisting of at least 8 nucleotides (*i.e.*, a hybridizable portion) of a human E-cadherin sequence; in other embodiments, the nucleic acids consist of at least 30 nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200 nucleotides of a human E-cadherin sequence. The invention also relates to nucleic acids hybridizable to or complementary to the foregoing sequences. In specific aspects, nucleic acids are provided which comprise a sequence complementary to at least 10, 25, 30, 50, 100, or 200 nucleotides or the entire coding region of a human E-cadherin gene. The longest stretch of identity among the human E-cadherin sequence of Figure 3 and the published mouse and chicken E-cadherin sequences is 29 nucleotides. Nucleic acids comprising a portion of the noncoding sequence shown in Figure 3 are also provided, as are nucleic acids complementary thereto. The nucleic acids of the invention do not consist of the nucleotide sequence shown in Figure 3 (SEQ ID NO:1) from nucleotide numbers 617-1036; preferably, such nucleic acids also do not consist of a portion of such nucleotide sequence.

Nucleic acids encoding fragments and derivatives of human E-cadherin proteins (see Section 5.6) are additionally provided.

In a preferred, but not limiting, aspect of the invention, a human E-cadherin DNA can be cloned and sequenced by the method described in Section 6, *infra*.

Any human cell potentially can serve as the nucleic acid source for the molecular cloning of the E-cadherin gene. The DNA may be obtained by standard procedures known in the art from cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell. (See, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985,

DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will lack introns and will contain only exon sequences. Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNAse in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, if an amount of a portion of a E-cadherin (of any species) gene or its specific RNA, or a fragment thereof, *e.g.*, an extracellular, or cytoplasmic region (see Section 5.6), is available and can be purified, or synthesized, and labeled, the generated DNA fragments may be screened by nucleic acid hybridization to the labeled probe (Benton and Davis, 1977, Science 196:180; Grunstein and Hogness, 1975, Proc. Natl. Acad. Sci. U.S.A. 72:3961). Those DNA fragments with substantial homology to the probe will hybridize. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known restriction map, either available or deduced from a known nucleotide sequence. Further selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be selected which produce a protein that, *e.g.*, has similar or identical electrophoretic migration, isolectric focusing behavior, proteolytic

digestion maps, binding activity, or antigenic properties as known for a E-cadherin protein. By use of an antibody to a E-cadherin protein, the E-cadherin protein may be identified by binding of labeled antibody to the putatively E-cadherin protein synthesizing clones, in an ELISA (enzyme-linked immunosorbent assay)-type procedure.

The E-cadherin gene can also be identified by mRNA selection by nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified E-cadherin DNA of human or of another species (*e.g.*, mouse, chicken). Immunoprecipitation analysis or functional assays (*e.g.*, binding to a receptor or ligand; see *infra*) of the *in vitro* translation products of the isolated products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. In addition, specific mRNAs may be selected by adsorption of polysomes isolated from cells to immobilized antibodies specifically directed against a E-cadherin protein. A radiolabelled E-cadherin cDNA can be synthesized using the selected mRNA (from the adsorbed polysomes) as a template. The radiolabelled mRNA or cDNA may then be used as a probe to identify the E-cadherin DNA fragments from among other genomic DNA fragments.

Alternatives to isolating the human E-cadherin genomic DNA include, but are not limited to, chemically synthesizing the gene sequence itself from a known sequence or making cDNA to the mRNA which encodes a human E-cadherin protein. For example, RNA for cDNA cloning of the human E-cadherin gene can be isolated from human cells (*e.g.*, epithelial cells) which express a E-cadherin protein. Other methods are possible and within the scope of the invention.

The identified and isolated gene can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell

used. Such vectors include, but are not limited to, bacteriophages such as lambda derivatives, or plasmids such as PBR322 or pUC plasmid derivatives. The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and E-cadherin gene may be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via transformation, transfection, infection, microinjection, electroporation, etc., so that many copies of the gene sequence are generated.

In an alternative method, the desired gene may be identified and isolated after insertion into a suitable cloning vector in a "shot gun" approach. Enrichment for the desired gene, for example, by size fractionation, can be done before insertion into the cloning vector.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate the isolated E-cadherin gene, cDNA, or synthesized DNA sequence enables generation of multiple copies of the gene. Thus, the gene may be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

5.2. EXPRESSION OF THE HUMAN E-CADHERIN GENE

The nucleotide sequence coding for a human E-cadherin protein or a functionally active fragment or other derivative thereof (see Section 5.6), can be inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted protein-

coding sequence. The necessary transcriptional and translational signals can also be supplied by the native E-cadherin gene and/or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence. These include but are not limited to mammalian cell systems infected with virus (*e.g.*, vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (*e.g.*, baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used. In a specific embodiment, a chimeric protein comprising the extracellular domain or repeat region or other domain of a human E-cadherin protein is expressed. In other specific embodiments, a full-length human E-cadherin cDNA is expressed, or a sequence encoding a functionally active portion of a human E-cadherin protein. In yet another embodiment, a fragment of a human E-cadherin protein comprising a domain of the protein, or other derivative, or analog of a human E-cadherin protein is expressed.

Any of the methods previously described for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of a nucleic acid sequence encoding a human E-cadherin protein or peptide fragment may be regulated by a second nucleic acid sequence so that the E-cadherin protein or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of a E-cadherin protein may be controlled by any promoter/enhancer element known in the art. Promoters which may be used to control E-cadherin gene expression include, but are not limited to, the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3′ long terminal repeat of Rous sarcoma virus (Yamamoto et al., 1980, Cell 22:787-797),

the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42), an adenovirus promoter, cytomegalovirus (CMV) promoter); prokaryotic promoters such as the β-lactamase (Villa-Kamaroff et al., 1978, Proc. Natl. Acad. Sci. U.S.A. 75:3727-3731), *tac* (DeBoer et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:21-25), λP_L, or *trc* promoters; see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region or the cauliflower mosaic virus 35S RNA promoter (Gardner et al., 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38:639-646; Ornitz et al., 1986, Cold Spring Harbor Symp. Quant. Biol. 50:399-409; MacDonald, 1987, Hepatology 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, Cell 38:647-658; Adames et al., 1985, Nature 318:533-538; Alexander et al., 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648; Hammer et al., 1987, Science 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315:338-340; Kollias et al., 1986, Cell 46:89-94; myelin basic

protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378). In a specific embodiment, a pGEX vector (Pharmacia) is used for expression in bacteria. In another specific embodiment, a nucleotide sequence encoding a human E-cadherin or fragment or derivative thereof is operatively linked to a promoter, wherein the promoter is not a human E-cadherin gene promoter.

Expression vectors containing human E-cadherin gene inserts can be identified by three general approaches: (a) nucleic acid hybridization, (b) presence or absence of "marker" gene functions, and (c) expression of inserted sequences. In the first approach, the presence of a foreign gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted E-cadherin gene. In the second approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, transformation phenotype, occlusion body formation in baculovirus, etc.) caused by the insertion of foreign genes in the vector. For example, if the E-cadherin gene is inserted within the marker gene sequence of the vector, recombinants containing the E-cadherin insert can be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying the foreign gene product expressed by the recombinant. Such assays can be based, for example, on the physical or functional properties of the E-cadherin gene product in *in vitro* assay systems, *e.g.*, binding to a ligand or receptor, binding with antibody.

Once a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As previously

explained, the expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (*e.g.*, lambda), and plasmid and cosmid DNA vectors, to name but a few.

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered E-cadherin protein may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.*, glycosylation, cleavage) of proteins. For example, mammalian, yeast, and baculovirus host cells can glycosylate proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed.

Both cDNA and genomic sequences can be cloned and expressed.

5.3. IDENTIFICATION AND PURIFICATION OF THE EXPRESSED GENE PRODUCTS

Once a recombinant which expresses a human E-cadherin gene sequence is identified, the gene product can be analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labelling of the product followed by analysis by gel electrophoresis, immunoassay, etc.

Once a human E-cadherin protein is identified, it may be isolated and purified by standard methods including chromatography (*e.g.*, ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. The functional properties may be evaluated using any suitable assay (see Section 5.7).

Alternatively, the amino acid sequence of a human E-cadherin protein can be deduced from the nucleotide sequence of the chimeric gene

contained in the recombinant. Once the amino acid sequence is thus known, the protein can be synthesized by standard chemical methods known in the art (*e.g.*, see Hunkapiller et al., 1984, Nature 310:105-111).

By way of example, the deduced amino acid sequence (SEQ ID NO:2) of a human E-cadherin protein is presented in Figure 3. In a specific embodiment of the present invention, a human E-cadherin protein, whether produced by recombinant DNA techniques or by chemical synthetic methods, includes but is not limited to one containing, as a primary amino acid sequence, all or part of the amino acid sequence substantially as depicted in Figure 3 (SEQ ID NO:2), as well as fragments and other derivatives, and analogs thereof. In specific embodiments, the invention relates to mature human E-cadherin proteins, *e.g.*, those having an amino acid sequence substantially as depicted in Figure 3 from amino acid numbers 151-878. In another specific embodiment, a protein comprises the amino acid sequence as depicted in Figure 3 from amino acid numbers 153-878. In another specific embodiment, the invention relates to an E-cadherin protein having an amino acid sequence substantially as depicted in Figure 3 from amino acid numbers 1-878. Purified proteins comprising the foregoing sequences are also provided. In another specific embodiment, the invention provides purified human E-cadherin proteins and fragments thereof that are free of detergents, substantially non-denatured, and/or free of other human cell membrane components. In another embodiment, the human E-cadherin protein or fragment thereof (*e.g.*, comprising the extracellular domain) is glycosylated (*e.g.*, as obtained by expression in mammalian cells). In another embodiment, the human E-cadherin protein or fragment thereof is nonglycosylated (*e.g.*, as obtained by expression in bacteria). Nonglycosylated mature E-cadherin proteins are believed to be capable of homotypic binding.

5.4. STRUCTURE OF THE HUMAN E-CADHERIN GENE AND PROTEIN

The structure of the human E-cadherin gene and protein can be analyzed by various methods known in the art.

5.4.1. GENETIC ANALYSIS

The cloned DNA or cDNA corresponding to the E-cadherin gene can be analyzed by methods including but not limited to Southern hybridization (Southern, 1975, J. Mol. Biol. 98:503-517), Northern hybridization (see *e.g.*, Freeman et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:4094-4098), restriction endonuclease mapping (Maniatis, 1982, Molecular Cloning, A Laboratory, Cold Spring Harbor, New York), and DNA sequence analysis (see *infra*). Polymerase chain reaction (PCR; U.S. Patent Nos. 4,683,202, 4,683,195 and 4,889,818; Gyllenstein et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7652-7656; Ochman et al., 1988, Genetics 120:621-623; Loh et al., 1989, Science 243:217-220) followed by Southern hybridization with a E-cadherin-specific probe can allow the detection of the human E-cadherin gene in DNA from various cell types. Northern hybridization analysis can be used to determine the expression of the E-cadherin gene. Various cell types, at various states of development, differentiation, or activity can be tested for E-cadherin gene expression. The stringency of the hybridization conditions for both Southern and Northern hybridization can be manipulated to ensure detection of nucleic acids with the desired degree of relatedness to the specific E-cadherin probe used, whether it be human or other species.

Restriction endonuclease mapping can be used to roughly determine the genetic structure of the human E-cadherin gene. Restriction maps derived by restriction endonuclease cleavage can be confirmed by DNA sequence analysis. Alternatively, restriction maps can be deduced, once the nucleotide sequence is known.

DNA sequence analysis can be performed by any techniques known in the art, including but not limited to the method of Maxam and Gilbert (1980, Meth. Enzymol. 65:499-560), the Sanger dideoxy method (Sanger et al., 1977, Proc. Natl. Acad. Sci. U.S.A. 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699; Sequenase, U.S. Biochemical Corp.), or Taq polymerase, or use of an automated DNA sequenator (*e.g.*, Applied Biosystems, Foster City, CA). The cDNA sequence of a human

E-cadherin gene comprises the sequence substantially as depicted in Figure 3 (SEQ ID NO:1), and described in Section 6, *infra*.

5.4.2. PROTEIN ANALYSIS

The amino acid sequence of a human E-cadherin protein can be derived by deduction from the DNA sequence, or alternatively, by direct sequencing of the protein, *e.g.*, with an automated amino acid sequencer. The amino acid sequence of a representative human E-cadherin protein comprises the amino acid sequence substantially as depicted in Figure 3 (SEQ ID NO:2). In a specific embodiment, the sequence of the mature E-cadherin protein is substantially as depicted in Figure 3 from amino acid numbers 151-878. Comparison of the human sequence to other known sequences allows identification of functional domains within the molecule, including but not limited to the extracellular domain, transmembrane region, cytoplasmic domain, amino-terminal processed region, homotypic binding domain, repeat region, repeat #1, repeat #2, and repeat #3 (see also Section 7 *infra*).

The E-cadherin protein sequence can be further characterized by a hydrophilicity analysis (Hopp and Woods, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:3824). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of a E-cadherin protein and the corresponding regions of the gene sequence which encode such regions.

Secondary, structural analysis (Chou and Fasman, 1974, Biochemistry 13:222) can also be done, to identify regions of a E-cadherin protein that assume specific secondary structures.

Manipulation, translation, and secondary structure prediction, as well as open reading frame prediction and plotting, can also be accomplished using computer software programs available in the art.

Other methods of structural analysis can also be employed. These include but are not limited to X-ray crystallography (Engstom, 1974, Biochem. Exp. Biol. 11:7-13) and computer modeling (Fletterick and Zoller (eds.), 1986, Computer Graphics and Molecular Modeling, in Current Communications in

Molecular Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

5.5. GENERATION OF ANTIBODIES TO THE HUMAN E-CADHERIN PROTEIN AND DERIVATIVES THEREOF

According to the invention, a human E-cadherin protein, its fragments or other derivatives, or analogs thereof, may be used as an immunogen to generate antibodies which recognize such an immunogen. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In a preferred embodiment, antibodies which specifically bind to human E-cadherin proteins are produced. In one embodiment, such an antibody recognizes the human E-cadherin protein having the sequence shown in Figure 3 (SEQ ID NO:2), or a portion thereof. In another embodiment, such an antibody specifically binds to human, but not mouse or chicken, E-cadherin. In another embodiment, antibodies to a particular domain (*e.g.*, the cytoplasmic domain) of a human E-cadherin protein are produced.

Various procedures known in the art may be used for the production of polyclonal antibodies to a human E-cadherin protein or derivative or analog. For the production of antibody, various host animals can be immunized by injection with a native human E-cadherin protein, or a synthetic version, or derivative (*e.g.*, fragment) thereof, including but not limited to rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum.

In a preferred embodiment, polyclonal or monoclonal antibodies are produced by use of a hydrophilic portion of a human E-cadherin peptide (*e.g.*, identified by the procedure of Hopp and Woods (1981, Proc. Natl. Acad. Sci. U.S.A. 78:3824)).

For preparation of monoclonal antibodies directed toward a E-cadherin protein sequence or analog thereof, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) can be used. In an additional embodiment of the invention, monoclonal antibodies can be produced in germ-free animals (PCT Publication No. WO 89/12690 dated December 28, 1989). According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030) or by transforming human B cells with EBV virus *in vitro* (Cole et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, pp. 77-96), or by other methods known in the art. In fact, according to the invention, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci. U.S.A. 81:6851-6855; Neuberger et al., 1984, Nature 312:604-608; Takeda et al., 1985, Nature 314:452-454) by splicing the genes from a mouse antibody molecule specific for a E-cadherin protein together with genes from a human antibody molecule of appropriate biological activity can be used; such antibodies are within the scope of this invention. Also within the scope of the invention are "humanized" antibodies (see, *e.g.*, EP Publication 239,400 dated September 30, 1987 by Winter).

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce E-cadherin protein-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for E-cadherin proteins, derivatives, or analogs.

Antibody fragments which contain the idiotype (binding domain) of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragment, and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, *e.g.* ELISA (enzyme-linked immunosorbent assay). For example, to select antibodies which recognize a specific domain of a E-cadherin protein, one may assay generated hybridomas for a product which binds to a E-cadherin fragment containing such domain. For selection of an antibody specific to a human E-cadherin protein and not E-cadherin of another species (*e.g.*, mouse, chicken), one can select on the basis of positive binding to a human E-cadherin protein and a lack of binding to the E-cadherin protein of the other species.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of the protein sequences of the invention (*e.g.*, see Section 5.7, *infra*), *e.g.*, for imaging these proteins, measuring levels thereof in appropriate physiological samples, etc., diagnostically, and therapeutically, *e.g.*, for inhibiting E-cadherin function.

5.6. HUMAN E-CADHERIN PROTEIN DERIVATIVES AND ANALOGS

The invention further relates to derivatives (including but not limited to fragments) and analogs of human E-cadherin proteins.

The production and use of derivatives and analogs related to human E-cadherin proteins are within the scope of the present invention. In a specific embodiment, the derivative or analog is functionally active, *i.e.*, capable of exhibiting one or more functional activities associated with a full-length, wild-type human E-cadherin protein. As one example, such derivatives or analogs which have the desired immunogenicity or antigenicity can be used, for example,

in immunoassays, for immunization, for promotion or inhibition of E-cadherin protein activity, etc. Such molecules which retain, or alternatively inhibit, a desired human E-cadherin protein property, *e.g.*, binding to a receptor or ligand, such as possibly Notch protein, can be used as inducers, or inhibitors, respectively, of such property and its physiological correlates. Derivatives or analogs of E-cadherin proteins can be tested for the desired activity by procedures known in the art, including but not limited to the assays described in Section 5.7.

In particular, E-cadherin derivatives can be made by altering E-cadherin sequences by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences which encode substantially the same amino acid sequence as a human E-cadherin gene may be used in the practice of the present invention. These include but are not limited to nucleotide sequences comprising all or portions of human E-cadherin genes which are altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the E-cadherin derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of a human E-cadherin protein including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

In a specific embodiment of the invention, proteins consisting of or comprising a fragment of a human E-cadherin protein consisting of at least 30 amino acids of the E-cadherin protein is provided. In other embodiments, the fragment consists of at least 6, 10, 50, 75, or 100 amino acids of the E-cadherin protein. Another specific embodiment relates to a protein comprising a fragment of the amino acid sequence shown in Figure 3 (SEQ ID NO:2) from amino acid numbers 728-878, which can be bound by anti-E-cadherin antibody. In another specific embodiment of the invention, a purified protein is provided which comprises a derivative or fragment of a human E-cadherin protein, with the proviso that said purified protein is not a mature human E-cadherin protein comprising amino acids 151-878 as depicted in Figure 3 (SEQ ID NO:2).

The human E-cadherin protein derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned E-cadherin gene sequence can be modified by any of numerous strategies known in the art (Maniatis, 1990, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. In the production of the gene encoding a derivative or analog of a human E-cadherin protein, care should be taken to ensure that the modified gene remains within the same translational reading frame as the E-cadherin gene, uninterrupted by translational stop signals, in the gene region where the desired E-cadherin protein activity is encoded.

Additionally, the E-cadherin-encoding nucleic acid sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, including but not limited to, *in vitro* site-directed mutagenesis (Hutchinson et al., 1978, J. Biol. Chem 253:6551).

Manipulations of the human E-cadherin sequence may also be made at the protein level. Included within the scope of the invention are human E-cadherin protein fragments or other derivatives or analogs which are differentially modified during or after translation, *e.g.*, by acetylation, glycosylation or deglycosylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$, acetylation, formylation, oxidation, reduction, etc.

In addition, analogs and derivatives of human E-cadherin proteins can be chemically synthesized. For example, a peptide corresponding to a portion of a E-cadherin protein which comprises the desired domain (see Section 5.6.1), or which mediates the desired activity *in vitro* or *in vivo*, can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the human E-cadherin protein sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids, α-amino isobutyric acid, 4-aminobutyric acid, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, designer amino acids such as β-methyl amino acids, Cα-methyl amino acids, and Nα-methyl amino acids.

In a specific embodiment, the human E-cadherin derivative is a chimeric, or fusion, protein comprising a human E-cadherin protein or fragment thereof (preferably consisting of at least a domain or region of the E-cadherin protein, or at least 30 amino acids of the E-cadherin protein) joined at its amino or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. In one embodiment, such a chimeric protein is produced by recombinant expression of a nucleic acid encoding the protein (comprising a human E-cadherin-coding sequence joined in-frame to a coding sequence for a different protein). Such a chimeric product can be made by ligating the appropriate nucleic

acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer. A specific embodiment relates to a chimeric protein comprising a fragment of a E-cadherin protein which comprises a domain or motif of the E-cadherin protein, *e.g.*, the extracellular domain, transmembrane region, cytoplasmic domain, amino-terminal processed region, homotypic binding domain, conserved cysteine domain, HAV sequence, repeat region, repeat #1, repeat #2, repeat #3, or any combination of the foregoing (see Section 7). Another specific embodiment relates to a chimeric protein comprising a fragment of a human E-cadherin protein of at least six amino acids. In specific embodiments, the fusion protein comprises or consists of the extracellular portion of human E-cadherin or a functional fragment thereof joined via a peptide bond to a transmembrane domain joined via a peptide bond to the cytoplasmic (signalling) domain or functional fragment thereof of another receptor or adhesion molecule (*e.g.*, DCC (Deleted in Colorectal Cancer), P-cadherin, N-cadherin; N-CAM (neural cell adhesion molecule); receptor tyrosine kinases such as growth factor receptors like epidermal growth factor receptor, fibroblast growth factor receptor, neu, etc.). Particular examples of human E-cadherin fusion proteins, consisting of a human E-cadherin fragment capable of generating anti-E-cadherin antibody fused to the carboxyl-terminus of glutathione-S-transferase, are described in Section 8 hereof. Another specific embodiment relates to a protein comprising portions of the human E-cadherin sequence which appear in different order or are missing amino acid sequence relative to native E-cadherin.

In another specific embodiment, a protein comprising a portion of the E-cadherin amino acid sequence shown in Figure 3 (SEQ ID NO:2) is provided, with the proviso that the protein does not contain amino acids numbers 153-307.

Other specific embodiments of derivatives and analogs are described in the subsections below and examples sections *infra*.

5.6.1. DERIVATIVES OF THE HUMAN E-CADHERIN PROTEIN CONTAINING ONE OR MORE DOMAINS OF THE PROTEIN

In a specific embodiment, the invention relates to human E-cadherin protein derivatives and analogs, in particular human E-cadherin fragments and derivatives of such fragments, that comprise one or more domains of a human E-cadherin protein, including but not limited to the extracellular domain, transmembrane region, cytoplasmic domain, amino-terminal processed region, homotypic binding domain, HAV sequence, conserved cysteine domain, repeat domain, repeat #1, repeat #2, and repeat #3. The amino acid sequences representing the foregoing domains for the protein having the sequence shown in Figure 3 are described in Section 7 *infra*. In a specific embodiment, a protein comprises one or more of human E-cadherin repeats #1, #2, and #3. In another specific embodiment, a protein comprises the human E-cadherin transmembrane region and cytoplasmic domain.

In another specific embodiment, the invention relates to a derivative or analog of human E-cadherin that lacks one or more domains of a human E-cadherin protein.

5.6.2. DERIVATIVES OF THE HUMAN E-CADHERIN PROTEIN THAT MEDIATE BINDING TO PROTEINS

The invention also provides human E-cadherin fragments, and analogs or derivatives of such fragments, which mediate binding to other proteins, and nucleic acid sequences encoding the foregoing. In specific embodiments, such fragments, analogs, and derivatives which bind to alpha, beta or gamma catenin, actin, spectrin/fodrin, and/or ankyrin are envisioned.

5.7. ASSAYS OF HUMAN E-CADHERIN PROTEINS, DERIVATIVES AND ANALOGS

The functional activity of E-cadherin proteins, derivatives and analogs can be assayed by various methods known in the art.

For example, in one embodiment, where one is assaying for the ability to bind or compete with a wild-type human E-cadherin protein for binding

to anti-E-cadherin protein antibody, various immunoassays known in the art can be used, including but not limited to competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labelled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

The ability to bind to another protein (be it a second E-cadherin protein; alpha, beta, or gamma catenin; actin, spectrin/fodrin, ankyrin, 220 kD undercoat protein (Itoh et al., 1991, J. Cell Biol. 115(5):1449-1462), or otherwise) can be demonstrated by *in vitro* binding assays, noncompetitive or competitive, by methods known in the art. In another embodiment, the ability of an E-cadherin derivative or analog to bind to an identical molecule ("homotypic binding") or to native E-cadherin can be assayed by methods known in the art.

In another embodiment, physiological correlates of E-cadherin introduction into cells can be assayed. For example, the ability to suppress cell invasion can be assayed by known methods (see, *e.g.*, Vleminckx et al., 1991, Cell 66:107-119).

Other methods will be known to the skilled artisan and are within the scope of the invention.

5.8. THERAPEUTIC AND PROPHYLACTIC USES

The human E-cadherin proteins, derivatives (including fragments) and analogs thereof; antibodies thereto; nucleic acids encoding the human E-cadherin proteins, derivatives, and analogs, and E-cadherin antisense nucleic acids have therapeutic utility in the modulation of functions mediated by human E-cadherin. Such therapeutically useful molecules provided by the invention are termed herein "Therapeutics." The Therapeutics have therapeutic value for various diseases and disorders.

One specific embodiment of the invention relates to Therapeutics which antagonize, or inhibit, a E-cadherin protein function. Such Therapeutics are most preferably identified by use of known convenient *in vitro* assays, *e.g.*, based on their ability to inhibit binding of E-cadherin to other proteins, or inhibit any known E-cadherin function as assayed *in vitro*, although *in vivo* assays may also be employed. In a preferred embodiment, such a Therapeutic is a protein or derivative thereof comprising a functionally active fragment such as a fragment of a human E-cadherin protein which binds to another protein; such a Therapeutic can be used, *e.g.*, in soluble form, to competitively inhibit the function mediated by such fragment. In specific embodiments, a Therapeutic is a protein comprising the homotypic binding domain, or an analog/competitive inhibitor of a E-cadherin signal-transducing function, a nucleic acid capable of expressing one of the foregoing proteins, a human E-cadherin antisense nucleic acid (see *infra*), or an anti-human E-cadherin antibody which neutralizes a functional activity of E-cadherin.

In another embodiment of the invention, a nucleic acid containing a portion of a dysfunctional E-cadherin gene is used, to promote E-cadherin inactivation by homologous recombination (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

In another embodiment, Therapeutics can be used to promote E-cadherin function. Such Therapeutics include but are not limited to human E-cadherin proteins and derivatives and analogs of the invention which are functionally active, *i.e.*, they are capable of displaying one or more known

functional activities associated with a full-length (wild-type) E-cadherin protein. In a preferred aspect, such functional activity is the ability to suppress cell invasion or metastasis. In a specific embodiment, such a Therapeutic comprises one or more domains of the human E-cadherin protein, preferably the homotypic binding domain.

In a specific embodiment of the invention in which a Therapeutic which promotes E-cadherin function is introduced into or delivered into a cell which does not normally express E-cadherin or which is not an epithelial or human placental cell, the Therapeutic is a chimeric/fusion protein comprising (a) an extracellular domain or functional derivative thereof that is of human E-cadherin, (b) a transmembrane domain, and (c) a cytoplasmic signalling domain or functional derivative thereof of a protein normally expressed by a cell type representative of the cell to which the Therapeutic is delivered. For example, where the cell is a neural or mesenchymal cell, the cytoplasmic domain can be of N-CAM or N-cadherin. In a specific embodiment relating to gene therapy, a recombinant nucleic acid encoding and capable of expressing such a chimeric molecule is introduced into a host cell.

Further descriptions and sources of Therapeutics of the inventions are found in Sections 5.8.1 through 5.10.1 herein.

In a specific embodiment, a Therapeutic which antagonizes E-cadherin function is administered to promote cell invasion. In another specific embodiment, a Therapeutic which promotes E-cadherin function is administered to inhibit cell invasion and metastasis. Thus, for example, introduction into cell of a nucleic acid encoding a human E-cadherin or fragment thereof which mediates homotypic binding is therapeutically useful for promotion of adhesion of such cells and prevention of their invasion and metastasis. Where such cells are tumor cells, direct delivery of the nucleic acid to such tumor cells *in vivo* is envisaged. In a different embodiment of the invention, such cells to be used for introduction of the E-cadherin nucleic acid can be any cells to be administered *in vivo* for therapeutic effect; introduction into such cells of the E-cadherin-encoding sequences prior to administration of the cell to a patient can prevent the cell from

subsequently becoming or behaving like an invasive tumor cell (see also Section 5.9).

Thus, in a preferred aspect of the invention, a Therapeutic which exhibits E-cadherin homotypic binding ability, is administered to treat or prevent malignancy, or metastasis of a malignancy. This is described further in Sections 5.8.1 and 5.8.2.

In another embodiment of the invention, a Therapeutic which promotes E-cadherin function (*e.g.*, comprising the extracellular domain or homotypic binding domain, and thus capable of mediating homotypic binding and resultant adhesion of cells attached to or expressing the Therapeutic) is used for treatment of benign dysproliferative disorders. Specific embodiments are directed to treatment of cirrhosis of the liver (a condition in which scarring has overtaken normal liver regeneration processes), treatment of keloid (hypertrophic scar) formation (disfiguring of the skin in which the scarring process interferes with normal renewal), psoriasis (a common skin condition characterized by excessive proliferation of the skin and delay in proper cell fate determination), and baldness (a condition in which terminally differentiated hair follicles fail to function properly).

In another embodiment of the invention, a Therapeutic which promotes E-cadherin function (*e.g.*, comprising the extracellular domain or homotypic binding domain, and thus capable of mediating homotypic binding) is used to promote wound healing, including the treatment of burns, and to promote the re-epithelialization of the skin, mucosal surfaces, or cornea. In a specific embodiment, fibroblasts obtained from a patient are transfected *in vitro* with a nucleic acid encoding human E-cadherin or a derivative thereof capable of homotypic binding, in order to form a synthetic skin graft that, when applied to the site of a patient's wound, provides a protective autologous barrier to promote wound healing. Incorporation of human E-cadherin, or cells expressing the same, in synthetic organs is also envisioned.

In yet another embodiment, a Therapeutic which promotes E-cadherin function is used to treat or prevent gestational disease, or fetal

wastage, for example, spontaneous abortions, and developmental abnormalities of the fetus or neonate. In a preferred aspect, the Therapeutic is administered into the amniotic sac or intrauterinely. E-cadherin is normally expressed in human placenta.

In yet another embodiment, a Therapeutic which promotes E-cadherin function (in particular, its function in establishing an impermeability barrier in epithelial cells) can be used for the treatment or prevention of inflammatory disorders, *e.g.*, Crohn's disease or sclerosing cholangitis. Crohn's disease and sclerosing cholangitis are associated with decreased permeability of epithelial cells. It has been reported that E-cadherin is required to establish the impermeability barrier in epithelial cells.

5.8.1. MALIGNANCIES

Malignant and pre-neoplastic conditions which can be treated by administration of a Therapeutic, preferably one which promotes adhesiveness mediated by E-cadherin and thus exhibits E-cadherin homotypic binding ability, include but are not limited to those described below in this and the subsequent subsection.

Such malignancies and related disorders, include but are not limited to those listed in Table 1 (for a review of such disorders, see Fishman et al., 1985, *Medicine*, 2d Ed., J.B. Lippincott Co., Philadelphia):

TABLE 1
MALIGNANCIES AND RELATED DISORDERS

- Leukemia
  - acute leukemia
    - acute lymphocytic leukemia
    - acute myelocytic leukemia
      - myeloblastic
      - promyelocytic
      - myelomonocytic
      - monocytic
      - erythroleukemia
  - chronic leukemia
    - chronic myelocytic (granulocytic) leukemia
    - chronic lymphocytic leukemia
- Polycythemia vera
- Lymphoma
  - Hodgkin's disease
  - non-Hodgkin's disease
- Multiple myeloma
- Waldenström's macroglobulinemia
- Heavy chain disease
- Solid tumors
  - sarcomas and carcinomas
    - fibrosarcoma
    - myxosarcoma
    - liposarcoma
    - chondrosarcoma
    - osteogenic sarcoma
    - chordoma
    - angiosarcoma
    - endotheliosarcoma
    - lymphangiosarcoma
    - lymphangioendotheliosarcoma
    - synovioma
    - mesothelioma
    - Ewing's tumor
    - leiomyosarcoma
    - rhabdomyosarcoma
    - colon carcinoma
    - stomach cancer
    - pancreatic cancer
    - breast cancer
    - ovarian cancer

prostate cancer
squamous cell carcinoma
basal cell carcinoma
adenocarcinoma
sweat gland carcinoma
sebaceous gland carcinoma
papillary carcinoma
papillary adenocarcinomas
cystadenocarcinoma
medullary carcinoma
bronchogenic carcinoma
renal cell carcinoma
hepatoma
bile duct carcinoma
seminoma
Wilms' tumor
cervical cancer
testicular tumor
lung carcinoma
small cell lung carcinoma
bladder carcinoma
epithelial carcinoma
glioma
astrocytoma
medulloblastoma
craniopharyngioma
ependymoma
pinealoma
hemangioblastoma
acoustic neuroma
oligodendroglioma
menangioma
melanoma
neuroblastoma
retinoblastoma
germ cell neoplasm (teratocarcinoma, embryonal carcinoma, choriocarcinoma)
other gestational proliferative disease (*e.g.*, molar pregnancy)

In specific embodiments, malignancy or dysproliferative changes (such as metaplasias and dysplasias) are treated or prevented in epithelial tissues such as those in the cervix, esophagus, lung, breast, bladder, kidney, and colon.

5.8.2. PREVENTION OF MALIGNANCIES

The Therapeutics of the invention which exhibit adhesiveness to cells, or homotypic binding, can be administered to prevent progression to a neoplastic or malignant state, including but not limited to those disorders listed in Table 1. Such prophylactic use is indicated in conditions known or suspected of preceding progression to neoplasia or cancer, in particular, where non-neoplastic cell growth consisting of hyperplasia, metaplasia, or most particularly, dysplasia has occurred (for review of such abnormal growth conditions, see Robbins and Angell, 1976, *Basic Pathology*, 2d Ed., W.B. Saunders Co., Philadelphia, pp. 68-79.) Hyperplasia is a form of controlled cell proliferation involving an increase in cell number in a tissue or organ, without significant alteration in structure or function. As but one example, endometrial hyperplasia often precedes endometrial cancer. Metaplasia is a form of controlled cell growth in which one type of adult or fully differentiated cell substitutes for another type of adult cell. Metaplasia can occur in epithelial or connective tissue cells. Atypical metaplasia involves a somewhat disorderly metaplastic epithelium. Dysplasia is frequently a forerunner of cancer, and is found mainly in the epithelia; it is the most disorderly form of non-neoplastic cell growth, involving a loss in individual cell uniformity and in the architectural orientation of cells. Dysplastic cells often have abnormally large, deeply stained nuclei, and exhibit pleomorphism. Dysplasia characteristically occurs where there exists chronic irritation or inflammation, and is often found in the cervix, respiratory passages, oral cavity, and gall bladder.

Alternatively or in addition to the presence of abnormal cell growth characterized as hyperplasia, metaplasia, or dysplasia, the presence of one or more characteristics of a transformed phenotype, or of a malignant phenotype, displayed *in vivo* or displayed *in vitro* by a cell sample from a patient, can indicate the desirability of prophylactic/therapeutic administration of a Therapeutic of the invention. Such characteristics of a transformed phenotype include morphology changes, looser substratum attachment, loss of contact inhibition, loss of anchorage dependence, protease release, increased sugar transport, decreased

serum requirement, expression of fetal antigens, disappearance of the 250,000 dalton cell surface protein, etc. (see also *id.*, at pp. 84-90 for characteristics associated with a transformed or malignant phenotype).

In a specific embodiment, leukoplakia, a benign-appearing hyperplastic or dysplastic lesion of the epithelium, or Bowen's disease, a carcinoma *in situ*, are pre-neoplastic lesions indicative of the desirability of prophylactic intervention.

In another embodiment, fibrocystic disease (cystic hyperplasia, mammary dysplasia, particularly adenosis (benign epithelial hyperplasia), or atypical papillomatosis) is indicative of the desirability of prophylactic intervention.

In another embodiment, a patient with a strong family history of breast cancer is treated with a Therapeutic, for prevention of breast cancer.

In other embodiments, a patient which exhibits one or more of the following predisposing factors for malignancy is treated by administration of an effective amount of a Therapeutic: a chromosomal translocation associated with a malignancy (*e.g.*, the Philadelphia chromosome for chronic myelogenous leukemia, t(14;18) for follicular lymphoma, etc.), familial polyposis or Gardner's syndrome (possible forerunners of colon cancer), benign monoclonal gammopathy (a possible forerunner of multiple myeloma), and a first degree kinship with persons having a cancer or precancerous disease showing a Mendelian (genetic) inheritance pattern (*e.g.*, familial polyposis of the colon, Gardner's syndrome, hereditary exostosis, polyendocrine adenomatosis, medullary thyroid carcinoma with amyloid production and pheochromocytoma, Peutz-Jeghers syndrome, neurofibromatosis of Von Recklinghausen, retinoblastoma, carotid body tumor, cutaneous melanocarcinoma, intraocular melanocarcinoma, xeroderma pigmentosum, ataxia telangiectasia, Chediak-Higashi syndrome, albinism, Fanconi's aplastic anemia, and Bloom's syndrome; *see* Robbins and Angell, 1976, *Basic Pathology*, 2d Ed., W.B. Saunders Co., Philadelphia, pp. 112-113); or one of the foregoing precancerous conditions, or neurofibromatosis (*e.g.*, tuberous sclerosis, von Hippel-Lindau disease, multiple exostoses), genodermatosis (*e.g.*,

polydysplastic epidermolysis bullosa), immune deficiency syndrome (*e.g.*, Wiskott-Aldrich syndrome, X-linked agammaglobulinemia), chromosome breakage or polyploidy (*e.g.*, Down's syndrome) (see The Merck Manual of Diagnosis and Therapy, 1987, 15th Ed., Berkow et al. (eds.), Merck Sharp & Dohme Research Laboratories, NJ, p. 1207), etc.)

5.8.3. NERVOUS SYSTEM DISORDERS

In another embodiment of the invention, Therapeutics which antagonize E-cadherin function, and thus promote cell invasiveness, can be used therapeutically, *e.g.*, to promote tissue repair and regeneration. A particular embodiment is directed to the promotion of nerve regeneration. Such Therapeutics which antagonize E-cadherin function include but are not limited to human E-cadherin antisense nucleic acids, anti-human E-cadherin monoclonal antibodies (*e.g.*, directed against the homotypic binding region, repeat region), and human E-cadherin peptide fragments or analogs thereof (*e.g.* of the E-cadherin extracellular domain, which when administered preferably in soluble form will competitively inhibit homotypic binding or adhesiveness of E-cadherin).

Nervous system disorders which are thus envisioned for treatment include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death,

including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) malignant lesions, in which a portion of the nervous system is destroyed or injured by malignant tissue which is either a nervous system associated malignancy or a malignancy derived from non-nervous system tissue;

(iv) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(v) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(vi) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to. vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vii) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(viii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(ix) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting neurite extension or survival or differentiation of neurons. For example, and not by way of limitation, Therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or *in vivo*;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, etc., depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In a specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, or degenerative disease that may affect motor neurons as well as other components of the nervous

system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

Other therapeutic and prophylactic methods provided by the invention are described in Sections 5.9 through 5.10.1 *infra*.

5.9. GENE THERAPY

Nucleic acids encoding human E-cadherin or functional derivatives thereof, and expression vectors comprising the same can be introduced into cells such that the E-cadherin nucleic acid sequences are stably incorporated in the cell and capable of expression by the cell and/or its progeny cells. Such introduction can occur *in vivo*, by or following *in vivo* administration of the E-cadherin encoding nucleic acid; or *in vitro*, after which the recombinant cell can be introduced into an animal, most preferably a human, for purposes of gene therapy (*i.e.*, therapeutic benefit via expression of the protein encoded by the introduced, heterologous gene sequence) (for reviews relating to gene therapy, see, *e.g.*, Karson et al., 1992, J. Reprod. Med. 37(6):508-514; Thompson, 1992, Science 258:744-746; Cline, 1985, Pharmac. Ther. 29:69-92). In a preferred embodiment, a nucleic acid encoding the complete human E-cadherin protein or a functional derivative thereof capable of homotypic binding, is introduced into a cell of an animal, either *in vitro* followed by introduction of the transformed cell or *in vivo* (*e.g.*, by direct injection into the animal) to prevent progression to malignancy, to prevent cell invasion, or to prevent metastasis. In a specific embodiment, the nucleic acid is directly injected into or otherwise delivered to a tumor cell, to prevent metastasis. Such tumor cells include but are not limited to the solid tumors listed in Table I, *supra*. In an alternative embodiment, recombinant cells engineered to secrete an E-cadherin protein or derivative can be

used to provide soluble E-cadherin to competitively inhibit E-cadherin adhesive function on cells, thus promoting cell invasion.

Cells into which an E-cadherin-encoding nucleic acid can be introduced for purposes of gene therapy encompass any desired, available cell type, and include but are not limited to epithelial cells, endothelial cells, keratinocytes, fibroblasts, muscle cells, hepatocytes; blood cells such as T lymphocytes, B lymphocytes, monocytes, macrophages, neutrophils, eosinophils, megakaryocytes, granulocytes; various stem or progenitor cells, in particular hematopoietic stem or progenitor cells, *e.g.*, as obtained from bone marrow, umbilical cord blood, peripheral blood, fetal liver, etc. In a specific embodiment in which a tumor is treated by gene therapy, the E-cadherin-encoding nucleic acid, if not directly introduced into the tumor cell *in vivo*, is preferably introduced into a cell of the same cell type as the tumor cell.

In a preferred embodiment, the cell used for gene therapy is autologous to the patient.

In an embodiment in which cells are obtained, a nucleic acid encoding E-cadherin or a derivative thereof is introduced into the cells such that it is expressible by the cells or their progeny, and the recombinant cells are then administered *in vivo* for therapeutic effect, stem cells are preferred for use. Any stem cells which can be isolated and maintained *in vitro* can potentially be used in accordance with this embodiment of the present invention. Such stem cells include but are not limited to hematopoietic stem cells (HSC), stem cells of epithelial tissues such as the skin and the lining of the gut, and embryonic heart muscle cells. Epithelial stem cells are preferred for use.

Epithelial stem cells (ESCs) or keratinocytes can be obtained from tissues such as the skin and the lining of the gut by known procedures (Rheinwald, 1980, Meth. Cell Bio. 21A:229). In stratified epithelial tissue such as the skin, renewal occurs by mitosis of stem cells within the germinal layer, the layer closest to the basal lamina. Stem cells within the lining of the gut provide for a rapid renewal rate of this tissue. ESCs or keratinocytes obtained from the skin or lining of the gut of a patient or donor can be grown in tissue culture

(Rheinwald, 1980, Meth. Cell Bio. 21A:229; Pittelkow and Scott, 1986, Mayo Clinic Proc. 61:771). If the ESCs are provided by a donor, a method for suppression of host versus graft reactivity (*e.g.*, irradiation, drug or antibody administration to promote moderate immunosuppression) can also be used.

With respect to hematopoietic stem cells (HSC), any technique which provides for the isolation, propagation, and maintenance *in vitro* of HSC can be used in this embodiment of the invention. Techniques by which this may be accomplished include (a) the isolation and establishment of HSC cultures from bone marrow cells isolated from the future host, or a donor, or (b) the use of previously established long-term HSC cultures, which may be allogeneic or xenogeneic. Non-autologous HSC are used preferably in conjunction with a method of suppressing transplantation immune reactions of the future host/patient. In a particular embodiment of the present invention, human bone marrow cells can be obtained from the posterior iliac crest by needle aspiration (see, *e.g.*, Kodo et al., 1984, J. Clin. Invest. 73:1377-1384). In a preferred embodiment of the present invention, the HSCs can be made highly enriched or in substantially pure form. This enrichment can be accomplished before, during, or after long-term culturing, and can be done by any techniques known in the art. Long-term cultures of bone marrow cells can be established and maintained by using, for example, modified Dexter cell culture techniques (Dexter et al., 1977, J. Cell Physiol. 91:335) or Witlock-Witte culture techniques (Witlock and Witte, 1982, Proc. Natl. Acad. Sci. USA 79:3608-3612).

In one embodiment, the nucleic acid encoding E-cadherin or a derivative thereof is introduced into a cell prior to administration *in vivo* of the resulting recombinant cell. Such introduction can be carried out by any method known in the art, including but not limited to transfection, electroporation, microinjection, infection with a viral or bacteriophage vector containing the E-cadherin sequences, cell fusion, chromosome-mediated gene transfer, microcell-mediated gene transfer, spheroplast fusion, etc. Numerous techniques are known in the art for the introduction of foreign genes into cells (see *e.g.*, Cline, 1985, Pharmac. Ther. 29:69-92) and may be used in accordance with the present

invention, provided that the necessary developmental and physiological functions of the recipient cells are not disrupted. The technique should provide for the stable transfer of the heterologous gene sequence to the cell, so that the heterologous gene sequence is expressible by the cell and preferably heritable and expressible by its cell progeny.

The resulting recombinant cells can be introduced by various methods known in the art (see Section 5.12 *infra*). In a preferred embodiment, epithelial cells are injected, *e.g.*, subcutaneously. In another embodiment, recombinant skin cells may be applied as a skin graft onto the patient. Recombinant blood cells (*e.g.*, hematopoietic stem or progenitor cells) are preferably administered intravenously. The amount of cells envisioned for use depends on the desired effect, patient state, etc., and can be determined by one skilled in the art.

In another specific embodiment, the nucleic acid encoding E-cadherin or a derivative thereof is directly administered *in vivo* for therapeutic effect, whereby it is expressed to produce E-cadherin or a derivative thereof. This can be accomplished by any of numerous methods known in the art, *e.g.*, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that it becomes intracellular, *e.g.*, by infection using a defective or attenuated retroviral or other viral vector (see U.S. Patent No. 4,980,286), or by direct injection, or by use of microparticle bombardment (*e.g.*, a gene gun; Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, or by administering it in linkage to a peptide which is known to enter the nucleus, etc. Alternatively, the nucleic acid Therapeutic can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination.

5.10. ANTISENSE REGULATION OF HUMAN E-CADHERIN EXPRESSION

The present invention also provides the therapeutic or prophylactic use of nucleic acids of at least six nucleotides that are antisense to a gene or cDNA encoding a human E-cadherin protein or a portion thereof. "Antisense" as

used herein refers to a nucleic acid capable of hybridizing to a portion of a human E-cadherin RNA (preferably mRNA) by virtue of some sequence complementarity. Such antisense nucleic acids have utility as antagonists of E-cadherin function, and can be used where cell invasion is desired (*e.g.*, to promote nerve or other tissue regeneration).

The antisense nucleic acids of the invention can be oligonucleotides that are double-stranded or single-stranded, RNA or DNA or a modification or derivative thereof, which can be directly administered to a cell, or which can be produced intracellularly by transcription of exogenous, introduced sequences.

The invention further provides pharmaceutical compositions comprising an effective amount of the E-cadherin antisense nucleic acids of the invention in a pharmaceutically acceptable carrier, as described in Section 5.12.

In another embodiment, the invention is directed to methods for inhibiting the expression of a human E-cadherin nucleic acid sequence in a prokaryotic or eukaryotic cell, comprising providing the cell with an effective amount of a composition comprising an antisense E-cadherin nucleic acid of the invention.

Human E-cadherin antisense nucleic acids and their uses are described in detail below.

5.10.1. HUMAN E-CADHERIN ANTISENSE NUCLEIC ACIDS

The human E-cadherin antisense nucleic acids are of at least six nucleotides and are preferably oligonucleotides (ranging from 6 to about 50 oligonucleotides). In specific aspects, the oligonucleotide is at least 10 nucleotides, at least 15 nucleotides, at least 30 nucleotides, at least 50 nucleotides, at least 100 nucleotides, or at least 200 nucleotides. In a specific embodiment, the oligonucleotide has a sequence that is not identical or 100% complementary to any same-size portion of the mouse or chicken E-cadherin nucleotide coding sequences or flanking sequences. In another specific embodiment, the antisense oligonucleotide is not 100% complementary to a same size nucleic acid having a sequence depicted in Figure 3 from nucleotide numbers 617-1036 or a portion

thereof. In another embodiment, the oligonucleotide is complementary to the nucleotide sequence encoding a domain or portion thereof of the human E-cadherin protein. The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone. The oligonucleotide may include other appending groups such as peptides, or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No. WO 88/09810, published December 15, 1988) or blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134, published April 25, 1988), hybridization-triggered cleavage agents (see, *e.g.*, Krol et al., 1988, BioTechniques 6:958-976) or intercalating agents (see, *e.g.*, Zon, 1988, Pharm. Res. 5:539-549). In a specific embodiment, the antisense nucleic acid is antisense to a sequence encoding one or more domains of the E-cadherin protein.

In a preferred aspect of the invention, a human E-cadherin antisense oligonucleotide is provided, preferably of single-stranded DNA. In a most preferred aspect, such an oligonucleotide comprises a sequence antisense to the sequence encoding the extracellular domain of a human E-cadherin protein, or the repeat domain thereof. The oligonucleotide may be modified at any position on its structure with substituents generally known in the art.

The E-cadherin antisense oligonucleotide may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil,

5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

In another embodiment, the oligonucleotide comprises at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the oligonucleotide is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641).

The oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

Oligonucleotides of the invention may be synthesized by standard methods known in the art, *e.g.* by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligos may be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), methylphosphonate oligos can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

In a specific embodiment, the E-cadherin antisense oligonucleotide comprises catalytic RNA, or a ribozyme (see, *e.g.*, PCT International Publication WO 90/11364, published October 4, 1990; Sarver et al., 1990, Science 247:1222-

1225). In another embodiment, the oligonucleotide is a 2'-0-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

In an alternative embodiment, the E-cadherin antisense nucleic acid of the invention is produced intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced *in vivo* such that it is taken up by a cell, within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid (RNA) of the invention. Such a vector would contain a sequence encoding the E-cadherin antisense nucleic acid. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the E-cadherin antisense RNA can be by any promoter known in the art to act in mammalian, preferably human, cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42), etc.

The antisense nucleic acids of the invention comprise a sequence complementary to at least a portion of an RNA transcript of a human E-cadherin gene. However, absolute complementarity, although preferred, is not required. A sequence "complementary to at least a portion of an RNA," as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded E-cadherin antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will

depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with a E-cadherin RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex. In another embodiment, 100% complementary sequences are envisioned.

The amount of E-cadherin antisense nucleic acid which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques.

E-cadherin antisense nucleic acids can be administered by methods as described *supra* in Section 5.9.

5.11. DEMONSTRATION OF THERAPEUTIC OR PROPHYLACTIC UTILITY

The Therapeutics of the invention can be tested *in vivo* for the desired therapeutic or prophylactic activity. For example, such compounds can be tested in suitable animal model systems prior to testing in humans, including but not limited to rats, mice, chicken, cows, monkeys, rabbits, etc. For *in vivo* testing, prior to administration to humans, any animal model system known in the art may be used.

5.12. THERAPEUTIC/PROPHYLACTIC ADMINISTRATION AND COMPOSITIONS

The invention provides methods of treatment (and prophylaxis) by administration to a subject of an effective amount of a Therapeutic of the invention. In a preferred aspect, the Therapeutic is substantially purified. The subject is preferably an animal, including but not limited to animals such as cows, pigs, chickens, etc., and is preferably a mammal, and most preferably human.

Various delivery systems are known and can be used to administer a Therapeutic of the invention, *e.g.*, encapsulation in liposomes, microparticles,

microcapsules, expression by recombinant cells, receptor-mediated endocytosis (see, *e.g.*, Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432), construction of a Therapeutic nucleic acid as part of a retroviral or other vector, etc. Methods of introduction include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, and oral routes. The compounds may be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal and intestinal mucosa, etc.) and may be administered together with other biologically active agents. Administration can be systemic or local. In addition, it may be desirable to introduce the pharmaceutical compositions of the invention into the central nervous system by any suitable route, including intraventricular and intrathecal injection; intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir. In a specific embodiment, it may be desirable to utilize liposomes targeted via antibodies to specific identifiable tumor antigens (Leonetti et al., 1990, Proc. Natl. Acad. Sci. U.S.A. 87:2448-2451; Renneisen et al., 1990, J. Biol. Chem. 265:16337-16342).

In a specific embodiment where the Therapeutic is a nucleic acid encoding a protein Therapeutic, the nucleic acid can be administered *in vivo* to promote expression of its encoded protein, by methods as described *supra* in Section 5.9.

In a specific embodiment, it may be desirable to administer the Therapeutics of the invention locally to the area in need of treatment; this may be achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers. In one embodiment, administration can be by direct injection at the site (or former site) of a malignant tumor or neoplastic or pre-neoplastic tissue.

The present invention also provides pharmaceutical compositions. Such compositions comprise a therapeutically effective amount of a Therapeutic, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The carrier and composition can be sterile. The formulation should suit the mode of administration.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc.

In a preferred embodiment, the composition is formulated in accordance with routine procedures as a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compositions for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The Therapeutics of the invention can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic,

tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The amount of the Therapeutic of the invention which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for intravenous administration of a protein Therapeutic are generally about 20-500 micrograms of active compound per kilogram body weight. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight to 1 mg/kg body weight. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

Suppositories generally contain active ingredient in the range of 0.5% to 10% by weight; oral formulations preferably contain 10% to 95% active ingredient.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

5.13. DIAGNOSTIC UTILITY

Detection and/or measurement of human E-cadherin expression has diagnostic and prognostic utility. The loss of expression or improper localization of expressed E-cadherin correlates with severity and degree of differentiation in

some cancers (see Shimoyama and Hirohashi, 1991, Cancer Res. 51:2185-2192). Thus, decreased expression or change in localization of E-cadherin in human tumor cells relative to the level of expression or localization, respectively, in non-malignant cells (preferably of the same cell type), indicates a poor prognosis and the presence of an invasive malignancy. Monitoring of the course of disease and of treatment efficacy can also be performed; such a decreased expression or change in localization relative to the level of expression or localization, respectively, in comparable cells taken from the patient at an earlier time (*e.g.*, a prior tissue biopsy sample taken, for example, prior to treatment) indicates the progression of malignancy or a poor response to treatment. In a specific embodiment, an anti-human E-cadherin antibody is used diagnostically in conventional immunoperoxidase staining of a surgical specimen to predict metastatic potential of an epithelial cell cancer such as breast, prostate, ovarian, gastric, or squamous cell cancer.

Disorders of cell fate, in particular precancerous conditions such as metaplasia and dysplasia, and hyperproliferative (*e.g.*, cancer) or hypoproliferative disorders, involving aberrant or undesirable levels of expression or activity of a E-cadherin protein can be diagnosed by detecting such levels. Thus, human E-cadherin proteins, analogues, derivatives, and subsequences thereof, E-cadherin nucleic acids (and sequences complementary thereto), anti-human-E-cadherin protein antibodies, and other proteins and derivatives and analogs thereof which interact with human E-cadherin proteins, and inhibitors of such E-cadherin-protein interactions, have uses in diagnostics. Such molecules can be used in assays, such as immunoassays, to detect, prognose, diagnose, or monitor various conditions, diseases, and disorders affecting E-cadherin expression, or monitor the treatment thereof. In particular, such an immunoassay is carried out by a method comprising contacting a sample derived from a patient with an anti-human E-cadherin protein antibody under conditions such that immunospecific binding can occur, and detecting or measuring the amount of any immunospecific binding by the antibody. In a specific embodiment, antibody to E-cadherin can be used to assay in a patient tissue or serum sample for the

presence of E-cadherin where an aberrant level of E-cadherin is an indication of a diseased condition.

The immunoassays which can be used include but are not limited to competitive and non-competitive assay systems using techniques such as western blots, radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, to name but a few.

Human E-cadherin genes and related nucleic acid sequences and subsequences, including complementary sequences, can also be used in hybridization assays. Human E-cadherin nucleic acid sequences, or subsequences thereof comprising about at least 8 nucleotides, can be used as hybridization probes. Hybridization assays can be used to detect, prognose, diagnose, or monitor conditions, disorders, or disease states associated with changes in human E-cadherin expression and/or activity as described *supra*. In particular, such a hybridization assay is carried out by a method comprising contacting a sample containing nucleic acid with a nucleic acid probe capable of hybridizing to human E-cadherin DNA or RNA, under conditions such that hybridization can occur, and detecting or measuring any resulting hybridization.

In another embodiment, PCR primers based on the human E-cadherin nucleotide sequence are used in diagnostic PCR. Thus, a pair of purified oligonucleotide primers is provided: a first oligonucleotide having a sequence which is the same as a first portion of the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1); and a second oligonucleotide having a sequence which is complementary to a second portion of the nucleotide sequence shown in Figure 3 (SEQ ID NO:1) and thus able to prime DNA synthesis off the opposite DNA strand from the first oligonucleotide; in which the second portion is situated 3′ to the first portion in the sequence shown in Figure 3. These oligonucleotides are then used as primers in PCR to amplify DNA fragments spanning from the first portion to the second portion in the E-cadherin gene, thus allowing the

detection of full-length E-cadherin nucleic acids or portions thereof (depending on the identity of the primers and thus their position within the E-cadherin gene) in a sample from a patient. The characteristics (*e.g.*, size) of the amplified fragment, or the lack of any amplification, where differing from that achieved with a wild-type human E-cadherin gene, can indicate the presence of an abnormality associated with a change in human E-cadherin gene sequence. In a specific embodiment directed to the PCR amplification of the complete coding sequence, primers from the noncoding regions, flanking the coding sequence, are employed. In specific embodiments, the oligonucleotide primers are at least 15, and are preferably 18 or 24 nucleotides. In particular, the primers consist of nucleotide sequences that are not identical or complementary to any same-size fragment of the chicken or mouse E-cadherin cDNAs or the coding regions or flanking regions thereof.

6. ISOLATION OF HUMAN E-CADHERIN cDNA CLONES FROM LIVER AND COLON

We have isolated and characterized the first full-length cDNA coding for human liver and colonic E-cadherin. The sequence is unique, although fairly homologous to similar isolates from mouse and chicken. We have isolated 10 cDNA clones from normal human liver and hepatocellular carcinoma libraries and over 20 cDNAs from a colonic epithelial cell cDNA library.

Our clones were obtained by hybridization screening of cDNA libraries received as gifts. We screened a normal colon cDNA library, a hepatocellular carcinoma library and a normal liver library. Our initial probes were made by polymerase chain reaction (PCR) from published human sequence (Mansouri et al., 1988, Differentiation 38:67-71), and from a fragment of human sequence (as published by Mansouri et al., *supra*) received as a gift from Dr. R. Kemler (see Fig. 2). These resulted in a single clone representing about 80% of the coding sequence. Further screening of the liver libraries and all screening of the colon library was done with restriction fragments from this or subsequent clones, until overlapping clones spanning the entire sequence were obtained.

7. THE COMPLETE NUCLEOTIDE AND AMINO ACID SEQUENCES OF HUMAN E-CADHERIN

Sequencing of human E-cadherin cDNA clones was carried out using Taq polymerase in dideoxy sequencing (TaqTrack, Promega, Madison, WI). The complete nucleotide sequence of the liver clone is shown schematically in Figure 2 with landmark regions mapped below the sequence. The 2756 bp human liver cDNA sequence predicts a 878 amino acid protein (Figure 3) which shows high levels of homology with the mouse (84% at the protein level and 80% at the DNA level) and chicken molecules. The sequence shows a 150 amino acid leader sequence (containing a signal sequence) that has 46% homology to (identity with) the mouse sequence, followed by the 26 amino acids representing the mature amino terminus that show 96% identity with the mouse sequence. The amino-terminal processed region (leader sequence) is cleaved post-translationally to produce the mature protein. The SHAVS sequence shown to be necessary for homotypic interaction is identical in sequence and location to that seen in mouse and chicken. Human E-cadherin shows the same three internal repeat (putative Ca^{++} binding) structures seen in mouse with 79-86% similarity to the mouse sequence and 22-36% similarity between repeats. The four cysteine residues in the region (conserved cysteine domain) immediately external to the transmembrane binding domain are identically conserved as are the three consensus sequences for N-glycosylation in that region in spite of the fact that this region is less similar overall (69% homology). The other N-glycosylation sequence in the middle of the second repeat is conserved but not identical. The region with the highest similarity is the carboxy terminal region of the molecule with the 24 transmembrane amino acids showing 100% identity and the cytoplasmic domain showing 95% similarity.

Thus, the human E-cadherin protein and domains thereof can be characterized as containing the following amino acids:

E-cadherin protein or portion thereof	Amino acids of Figure 3 (SEQ ID NO:2)
mature protein	151-878
extracellular domain	1-703; 151-703 of mature protein
transmembrane domain	704-727
cytoplasmic domain	728-878
amino-terminal processed region	1-150
homotypic binding domain	228-232
HAV homotypic binding sequence	229-231
conserved cysteine domain	513-703
Repeat domain	178-513
Repeat #1	178-289
Repeat #2	290-401
Repeat #3	402-513

8. cDNA CLONE SERIES, PROKARYOTIC EXPRESSION, AND ANTIBODY PRODUCTION

We used three individual clones to construct a coding sequence clone (termed FLEC) which encodes a protein consisting of 5 amino-terminal amino acids not from the human E-cadherin sequence, followed by the human E-cadherin protein sequence from approximately amino acid 146 through amino acid 878. The FLEC-encoded protein thus is missing amino acids 1 to about 145 of the sequence shown in Figure 3. The FLEC clone thus encodes the entire mature E-cadherin protein sequence (amino acids 151-878 of Fig. 3), but differs in the protein sequence corresponding to the amino-terminal processed region of human E-cadherin.

A full-length human E-cadherin coding sequence can be generated from bsFLEC and bsL5.1 (both deposited with the ATCC; see Section 10). Both bsFLEC and bsL5.1 represent E-cadherin sequences cloned in "bluescript phagemid II" from Stratgene Inc. L5.1 represents the 5' end of the E-cadherin clone (see Fig. 4) and is cloned into the EcoR1 site of the plasmid. bsFLEC is cloned between the EcoR1 and XhoI sites. To produce a full length human

E-cadherin construct, one cuts both plasmids with HindIII and BamHI and purifies the larger (4.7 kbp) band from bsFLEC and the 840 bp band from bsL5.1. These fragments are then ligated, and the resultant product is the full length clone. FLEC has been transferred into multiple eukaryotic cloning vectors.

We have also used the FLEC clone as a template to produce fragments for cloning into prokaryotic expression vectors (pGEX series, Pharmacia; pQE Series, Qiagen). Although detectable expression was not obtained with the pQE series, fusion proteins were obtained using the pGEX expression vector system in *E. coli* HB101. This vector system produces chimeric (fusion) proteins containing the carboxyl terminus of glutathione S-transferase as the amino-terminal sequence fused to the E-cadherin sequence. We have produced two fusion proteins, one, encoded by clone e250 (see Fig. 2), containing the extracellular domain including the homotypic adhesion sequence, and a second, encoded by clone cyto 20 (see Fig. 2) containing the entire cytoplasmic domain.

Clone e250 contains nucleotide numbers 589-832 of the human E-cadherin sequence (Fig. 2; SEQ ID NO:1). The e250-encoded protein thus contains the HAV binding domain, and is expected to competitively inhibit homotypic binding of E-cadherin. Clone e250 was made by subcloning a PvuII-BamHI fragment from clone H9.1 into pGEX-2T. Expression of clone e250 in *E. coli* HB101 yielded a 37 kd fusion protein as detected by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Clone cyto 20 contains nucleotide numbers 2297-2750 of the human E-cadherin sequence (Fig. 2; SEQ ID NO:1). Clone cyto 20 was made by performing PCR with 16-mer primers just 5′ and just 3′ of nucleotide numbers 2297-2749 in Figure 3. The primers were designed so as to incorporate an in-frame BamHI site at nucleotide number 2296. The PCR-amplified product was then cleaved with BamHI and subcloned into pGEX-3X. Expression in *E. coli* HB101 yielded a 42 kd fusion protein as detected by SDS-PAGE. Preliminary evidence indicates that the cyto 20-encoded fusion protein binds to brain

fodrin/spectrin, and co-precipates proteins out of MDCK cell extracts in the size range of alpha and beta catenin (see Ozawa and Kemler, 1992, J. Cell Biol. 116:989-96, regarding mouse E-cadherin).

We have produced both of the above-described fusion proteins in large quantity and generated polyclonal antibodies to both in rabbits. Fusion proteins were produced and purified by affinity chromatography according to standard procedures (Smith and Johnson, 1988, Gene 67:31-40) and utilized for immunization of rabbits (for polyclonal antibody production) and of mice (for monoclonal antibody production). The polyclonal antibodies thus obtained were tested against native human E-cadherin in colonic extracts and in extracts from a number of cell lines and were shown to bind with high affinity. We are currently in the process of producing mouse monoclonal antibodies to each fusion protein as follows: Since the fusion proteins appear to be toxic to mice, lymph nodes or spleen were taken for fusion with myeloma cells after only one immunization and one boost. The initial injection of fusion protein into mice was in Freund's complete adjuvant; the boost was given 13 days later in Freund's incomplete adjuvant. Three days after boosting, the lymph nodes or spleen were taken, and fusions with myeloma cells were performed, generating hybridomas for screening. The hybridomas were screened for reactivity with the fusion protein used as immunogen, and ten candidate hybridomas were identified as reactive with the e250-encoded protein, and are subject to verification.

We also have a large series of cloned fragments (Fig. 4) representing various portions of the molecule which may be used later for deletion analysis type studies.

9. EUKARYOTIC EXPRESSION OF HUMAN E-CADHERIN

We have subcloned the clone H9.1 DNA carboxy-terminal-encoding fragments (including the complete cytoplasmic domain) into eukaryotic expression vectors. We are using a series of vectors including, pCMV-NeoPoly I (kindly provided by E.Fearon, pers. comm; Fig. 6), pLXSN (Weintraub et al., 1989, Proc. Natl. Acad. Sci. USA 86:5434-5438), pGALO (Kato et al., 1990,

Mol. Cell Biol. 10:5914-5920), and pNLVP16 (Dang et al., 1991, Mol. Cell Biol. 11:954-962). Transfection experiments are underway with CHO (Chinese hamster ovary) cells, Jurkat (T lymphoma cell line) cells, and a number of breast cancer cell lines.

10. DEPOSIT OF MICROORGANISMS

The following bacterial strains, containing the listed plasmids, were deposited on November 12, 1992 with the American Type Culture Collection (ATCC), 1201 Parklawn Drive, Rockville, Maryland 20852, under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedures, and assigned the indicated accession numbers.

Bacteria	carrying	Plasmid	ATCC Accession Number
E. coli HB101		bsFLEC (containing FLEC)	69123
E. coli HB101		bsL5.1 (containing L5.1)	69122

The present invention is not to be limited in scope by the microorganisms described or the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Rimm, David L
Morrow, Jon S

(ii) TITLE OF INVENTION: Human Homolog of the E-Cadherin Gene and Methods Based Thereon.

(iii) NUMBER OF SEQUENCES: 2

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Pennie & Edmonds
(B) STREET: 1155 Avenue of the Americas
(C) CITY: New York
(D) STATE: New York
(E) COUNTRY: U.S.A.
(F) ZIP: 10036-2711

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: To be assigned
(B) FILING DATE: On even date herewith
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Misrock, S. Leslie
(B) REGISTRATION NUMBER: 18,872
(C) REFERENCE/DOCKET NUMBER: 7326-014

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (212)790-9090
(B) TELEFAX: (212)869-8864/9741
(C) TELEX: 66141 PENNIE

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2815 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 116..2749

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GAATTCCGGA AAGCACCTGT GAGCTTGGCA AGTCAGTTCA GAGCTCCAGC CCGCTCCAGC      60

CCGGCCCGAC CCGACCGCAC CCGGCGCCTG CCTCGCTCGG GCTCCCCGGC CAGCC ATG      118
                                                             Met
                                                               1

GGC CCT TGG AGC CGC AGC CTC TCG GGC CTG CTG CTG CTG CTG AGG TCT      166
Gly Pro Trp Ser Arg Ser Leu Ser Gly Leu Leu Leu Leu Leu Arg Ser
              5                  10                  15
```

```
CCT CTT GGC TCT CAG GAG CGG AGC CCT CCT CCC TGT TTG ACG CGA GAG     214
Pro Leu Gly Ser Gln Glu Arg Ser Pro Pro Pro Cys Leu Thr Arg Glu
         20                  25                  30

CTA CAC GTT CAC GGT GCC CCG GCG CCA CCT GAG AAG AGG CCG CGT CTG     262
Leu His Val His Gly Ala Pro Ala Pro Pro Glu Lys Arg Pro Arg Leu
     35                  40                  45

GGC AGA GTG AAT TTT GAA GAT TGC ACC GGT CGA CAA AGG ACA GCT ATT     310
Gly Arg Val Asn Phe Glu Asp Cys Thr Gly Arg Gln Arg Thr Ala Ile
 50                  55                  60                  65

TTC CTG ACA CCG ATT CCG AAA GTG GGC ACA GAT GGT GTG ATT ACA GTC     358
Phe Leu Thr Pro Ile Pro Lys Val Gly Thr Asp Gly Val Ile Thr Val
                 70                  75                  80

AAA AGG CCT CTA CGG TTT CAT AAC CCA ACA GAT CCA TTT CTT GGT CTA     406
Lys Arg Pro Leu Arg Phe His Asn Pro Thr Asp Pro Phe Leu Gly Leu
             85                  90                  95

CGC TGG GAC TCC ACC TAC AGA AAG TTT TCC ACC AAA GTC ACG CTG AAT     454
Arg Trp Asp Ser Thr Tyr Arg Lys Phe Ser Thr Lys Val Thr Leu Asn
        100                 105                 110

ACA GTG GGG CAC CAC CAC CGC CCC CCG CCC CAT CAG GCC TCC GTT TCT     502
Thr Val Gly His His His Arg Pro Pro Pro His Gln Ala Ser Val Ser
    115                 120                 125

GGA ATC CAA GCA GAA TTG CTC ACA TTT CCC AAC TCC TCT CCT GGC CTC     550
Gly Ile Gln Ala Glu Leu Leu Thr Phe Pro Asn Ser Ser Pro Gly Leu
130                 135                 140                 145

AGA AGA CAG AAG AGA GAC TGG GTT ATT CCT CCC ATC AGC TGC CCA GAA     598
Arg Arg Gln Lys Arg Asp Trp Val Ile Pro Pro Ile Ser Cys Pro Glu
                150                 155                 160

AAT GAA AAA GGC CCA TTT CCT AAA AAC CTG GTT CAG ATC AAA TCC AAC     646
Asn Glu Lys Gly Pro Phe Pro Lys Asn Leu Val Gln Ile Lys Ser Asn
            165                 170                 175

AAA GAC AAA GAA GGC AAG GTT TTC TAC AGC ATC ACT GGC CAA GGA GCT     694
Lys Asp Lys Glu Gly Lys Val Phe Tyr Ser Ile Thr Gly Gln Gly Ala
        180                 185                 190

GAC ACA CCC CCT GTT GGT GTC TTT ATT ATT GAA AGA GAA ACA GGA TGG     742
Asp Thr Pro Pro Val Gly Val Phe Ile Ile Glu Arg Glu Thr Gly Trp
    195                 200                 205

CTG AAG GTG ACA GAG CCT CTG GAT AGA GAA CGC ATT GCC ACA TAC ACT     790
Leu Lys Val Thr Glu Pro Leu Asp Arg Glu Arg Ile Ala Thr Tyr Thr
210                 215                 220                 225

CTC TTC TCT CAC GCT GTG TCA TCC AAC GGG AAT GCA GTT GAG GAT CCA     838
Leu Phe Ser His Ala Val Ser Ser Asn Gly Asn Ala Val Glu Asp Pro
                230                 235                 240

ATG GAG ATT TTG ATC ACG GTA ACC GAT CAG AAT GAC AAC AAG CCC GAA     886
Met Glu Ile Leu Ile Thr Val Thr Asp Gln Asn Asp Asn Lys Pro Glu
            245                 250                 255

TTC ACC CAG GAG GTC TTT AAG GGG TCT GTC ATG GAA GGT GCT CTT CCA     934
Phe Thr Gln Glu Val Phe Lys Gly Ser Val Met Glu Gly Ala Leu Pro
        260                 265                 270

GGA ACC TCT GTG ATG GAG GTC ACA GCC ACA GAC GCG GAC GAT GAT GTG     982
Gly Thr Ser Val Met Glu Val Thr Ala Thr Asp Ala Asp Asp Asp Val
    275                 280                 285

AAC ACC TAC AAT GCC GCC ATC GCT TAC ACC ATC CTC AGC CAA GAT CCT    1030
Asn Thr Tyr Asn Ala Ala Ile Ala Tyr Thr Ile Leu Ser Gln Asp Pro
```

```
290                 295                 300                 305

GAG CTC CCT GAC AAA AAT ATG TTC ACC ATT AAC AGG AAC ACA GGA GTC        1078
Glu Leu Pro Asp Lys Asn Met Phe Thr Ile Asn Arg Asn Thr Gly Val
                310                 315                 320

ATC AGT GTG GTC ACC ACT GGG CTG GAC CGA GAG AGT TTC CCT ACG TAT        1126
Ile Ser Val Val Thr Thr Gly Leu Asp Arg Glu Ser Phe Pro Thr Tyr
            325                 330                 335

ACC CTG GTG GTT CAA GCT GCT GAC CTT CAA GGT GAG GGG TTA AGC ACA        1174
Thr Leu Val Val Gln Ala Ala Asp Leu Gln Gly Glu Gly Leu Ser Thr
        340                 345                 350

ACA GCA ACA GCT GTG ATC ACA GTC ACT GAC ACC AAC GAT AAT CCT CCG        1222
Thr Ala Thr Ala Val Ile Thr Val Thr Asp Thr Asn Asp Asn Pro Pro
    355                 360                 365

ATC TTC AAT CCC ACC ACG TAC AAG GGT CAG GTG CCT GAG AAC GAG GCT        1270
Ile Phe Asn Pro Thr Thr Tyr Lys Gly Gln Val Pro Glu Asn Glu Ala
370                 375                 380                 385

AAC GTC GTA ATC ACC ACA CTG AAA GTG ACT GAT GCT GAT GCC CCC AAT        1318
Asn Val Val Ile Thr Thr Leu Lys Val Thr Asp Ala Asp Ala Pro Asn
                390                 395                 400

ACC CCA GCG TGG GAG GCT GTA TAC ACC ATA TTG AAT GAT GAT GGT GGA        1366
Thr Pro Ala Trp Glu Ala Val Tyr Thr Ile Leu Asn Asp Asp Gly Gly
            405                 410                 415

CAA TTT GTC GTC ACC ACA AAT CCA GTG AAC AAC GAT GGC ATT TTG AAA        1414
Gln Phe Val Val Thr Thr Asn Pro Val Asn Asn Asp Gly Ile Leu Lys
        420                 425                 430

ACA GCA AAG GGC TTG GAT TTT GAG GCC AAG CAG CAG TAC ATT CTA CAC        1462
Thr Ala Lys Gly Leu Asp Phe Glu Ala Lys Gln Gln Tyr Ile Leu His
    435                 440                 445

GTA GCA GTG ACG AAT GTG GTA CCT TTT GAG GTC TCT CTC ACC ACC TCC        1510
Val Ala Val Thr Asn Val Val Pro Phe Glu Val Ser Leu Thr Thr Ser
450                 455                 460                 465

ACA GCC ACC GTC ACC GTG GAT GTG CTG GAT GTG AAT GAA GGC CCC ATC        1558
Thr Ala Thr Val Thr Val Asp Val Leu Asp Val Asn Glu Gly Pro Ile
                470                 475                 480

TTT GTG CCT CCT GAA AAG AGA GTG GAA GTG TCC GAG GAC TTT GGC GTG        1606
Phe Val Pro Pro Glu Lys Arg Val Glu Val Ser Glu Asp Phe Gly Val
            485                 490                 495

GGC CAG GAA ATC ACA TCC TAC ACT GCC CAG GAG CCA GAC ACA TTT ATG        1654
Gly Gln Glu Ile Thr Ser Tyr Thr Ala Gln Glu Pro Asp Thr Phe Met
        500                 505                 510

GAA CAG AAA ATA ACA TAT CGG ATT TGG AGA GAC ACT CGC AAC TGG CTG        1702
Glu Gln Lys Ile Thr Tyr Arg Ile Trp Arg Asp Thr Arg Asn Trp Leu
    515                 520                 525

GAG ATT AAT CCG GAC ACT GGT GCC ATT TCC ACT CGG GCT GAG CTG GAC        1750
Glu Ile Asn Pro Asp Thr Gly Ala Ile Ser Thr Arg Ala Glu Leu Asp
530                 535                 540                 545

AGG GAG GAT TTT GAG CAC GTG AAG AAC AGC ACG TAC ACA GCC CTA ATC        1798
Arg Glu Asp Phe Glu His Val Lys Asn Ser Thr Tyr Thr Ala Leu Ile
                550                 555                 560

ATA GCT ACA GAC AAT GGT TCT CCA GTT GCT ACT GGA ACA GGG ACA CTT        1846
Ile Ala Thr Asp Asn Gly Ser Pro Val Ala Thr Gly Thr Gly Thr Leu
            565                 570                 575
```

CTG CTG ATC CTG TCT GAT GTG AAT GAC AAC GCC CCC ATA CCA GAA CCT 1894
Leu Leu Ile Leu Ser Asp Val Asn Asp Asn Ala Pro Ile Pro Glu Pro
580 585 590

CGA ACT ATA TTC TTC TGT GAG AGG AAT CCA AAG CCT CAG GTC ATA AAC 1942
Arg Thr Ile Phe Phe Cys Glu Arg Asn Pro Lys Pro Gln Val Ile Asn
595 600 605

ATT CAT GAT GCA GAC CTT CCT CCC AAT ACA TCT CCC TTC ACA GCA GAA 1990
Ile His Asp Ala Asp Leu Pro Pro Asn Thr Ser Pro Phe Thr Ala Glu
610 615 620 625

CTA ACA CAC GGG CGA GTG CCC AAC TGG ACC ATT CAG TAC AAC GAC CCA 2038
Leu Thr His Gly Arg Val Pro Asn Trp Thr Ile Gln Tyr Asn Asp Pro
630 635 640

ACC CAA GAA TCT ATC ATT TTG AAG CCA AAG ATG GCC TTA GAG GTG GGT 2086
Thr Gln Glu Ser Ile Ile Leu Lys Pro Lys Met Ala Leu Glu Val Gly
645 650 655

GAC TAC AAA ATC AAT CTC AAG CTC ATG GAT AAC CAG AAT AAA GAC CAA 2134
Asp Tyr Lys Ile Asn Leu Lys Leu Met Asp Asn Gln Asn Lys Asp Gln
660 665 670

GTG ACC ACC TTA GAG GTC AGC GTG TGT GAC TGT GAA GGG GCC GCC GGC 2182
Val Thr Thr Leu Glu Val Ser Val Cys Asp Cys Glu Gly Ala Ala Gly
675 680 685

GTC TGT AGG AAG GCA CAG CCT GTC GAA GCA GGA TTG CAA ATT CCT GCC 2230
Val Cys Arg Lys Ala Gln Pro Val Glu Ala Gly Leu Gln Ile Pro Ala
690 695 700 705

ATT CTG GGG ATT CTT GGA GGA ATT CTT GCT TTG CTA ATT CTG ATT CTG 2278
Ile Leu Gly Ile Leu Gly Gly Ile Leu Ala Leu Leu Ile Leu Ile Leu
710 715 720

CTG CTC TTG CTG TTT CTT CGG AGG AGA GCG GTG GTC AAA GAG CCC TTA 2326
Leu Leu Leu Leu Phe Leu Arg Arg Arg Ala Val Val Lys Glu Pro Leu
725 730 735

CTG CCC CCA GAG GAT GAC ACC CGG GAC AAC GTT TAT TAC TAT GAT GAA 2374
Leu Pro Pro Glu Asp Asp Thr Arg Asp Asn Val Tyr Tyr Tyr Asp Glu
740 745 750

GAA GGA GGC GGA GAA GAG GAC CAG GAC TTT GAC TTG AGC CAG CTG CAC 2422
Glu Gly Gly Gly Glu Glu Asp Gln Asp Phe Asp Leu Ser Gln Leu His
755 760 765

AGG GGC CTG GAC GCT CGG CCT GAA GTG ACT CGT AAC GAC GTT GCA CCA 2470
Arg Gly Leu Asp Ala Arg Pro Glu Val Thr Arg Asn Asp Val Ala Pro
770 775 780 785

ACC CTC ATG AGT GTC CCC CGG TAT CTT CCC CGC CCT GCC AAT CCC GAT 2518
Thr Leu Met Ser Val Pro Arg Tyr Leu Pro Arg Pro Ala Asn Pro Asp
790 795 800

GAA ATT GGA AAT TTT ATT GAT GAA AAT CTG AAA GCG GCT GAT ACT GAC 2566
Glu Ile Gly Asn Phe Ile Asp Glu Asn Leu Lys Ala Ala Asp Thr Asp
805 810 815

CCC ACA GCC CCG CCT TAT GAT TCT CTG CTC GTG TTT GAC TAT GAA GGA 2614
Pro Thr Ala Pro Pro Tyr Asp Ser Leu Leu Val Phe Asp Tyr Glu Gly
820 825 830

AGC GGT TCC GAA GCT GCT AGT CTG AGC TCC CTG AAC TCC TCA GAG TCA 2662
Ser Gly Ser Glu Ala Ala Ser Leu Ser Ser Leu Asn Ser Ser Glu Ser
835 840 845

GAC AAA GAC CAG GAC TAT GAC TAC TTG AAC GAA TGG GGC AAT CCG TTC 2710
Asp Lys Asp Gln Asp Tyr Asp Tyr Leu Asn Glu Trp Gly Asn Pro Phe

850 855 860 865

AAG AAG CTG GCT GAC ATG TAC GGA GGC GGC GAG GAC CAC TAGGGGACTC 2759
Lys Lys Leu Ala Asp Met Tyr Gly Gly Gly Glu Asp His
870 875

GAGAGAGGCG GCCCAGACCA TGTGCAGAAA TGCAGAAATC AGCGTTCTGT TGTTTT 2815

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 878 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gly Pro Trp Ser Arg Ser Leu Ser Gly Leu Leu Leu Leu Leu Arg
1 5 10 15

Ser Pro Leu Gly Ser Gln Glu Arg Ser Pro Pro Pro Cys Leu Thr Arg
20 25 30

Glu Leu His Val His Gly Ala Pro Ala Pro Pro Glu Lys Arg Pro Arg
35 40 45

Leu Gly Arg Val Asn Phe Glu Asp Cys Thr Gly Arg Gln Arg Thr Ala
50 55 60

Ile Phe Leu Thr Pro Ile Pro Lys Val Gly Thr Asp Gly Val Ile Thr
65 70 75 80

Val Lys Arg Pro Leu Arg Phe His Asn Pro Thr Asp Pro Phe Leu Gly
85 90 95

Leu Arg Trp Asp Ser Thr Tyr Arg Lys Phe Ser Thr Lys Val Thr Leu
100 105 110

Asn Thr Val Gly His His His Arg Pro Pro Pro His Gln Ala Ser Val
115 120 125

Ser Gly Ile Gln Ala Glu Leu Leu Thr Phe Pro Asn Ser Ser Pro Gly
130 135 140

Leu Arg Arg Gln Lys Arg Asp Trp Val Ile Pro Pro Ile Ser Cys Pro
145 150 155 160

Glu Asn Glu Lys Gly Pro Phe Pro Lys Asn Leu Val Gln Ile Lys Ser
165 170 175

Asn Lys Asp Lys Glu Gly Lys Val Phe Tyr Ser Ile Thr Gly Gln Gly
180 185 190

Ala Asp Thr Pro Pro Val Gly Val Phe Ile Ile Glu Arg Glu Thr Gly
195 200 205

Trp Leu Lys Val Thr Glu Pro Leu Asp Arg Glu Arg Ile Ala Thr Tyr
210 215 220

Thr Leu Phe Ser His Ala Val Ser Ser Asn Gly Asn Ala Val Glu Asp
225 230 235 240

Pro Met Glu Ile Leu Ile Thr Val Thr Asp Gln Asn Asp Asn Lys Pro
245 250 255

Glu Phe Thr Gln Glu Val Phe Lys Gly Ser Val Met Glu Gly Ala Leu
260 265 270

Pro Gly Thr Ser Val Met Glu Val Thr Ala Thr Asp Ala Asp Asp Asp
275 280 285

Val Asn Thr Tyr Asn Ala Ala Ile Ala Tyr Thr Ile Leu Ser Gln Asp
290 295 300

Pro Glu Leu Pro Asp Lys Asn Met Phe Thr Ile Asn Arg Asn Thr Gly
305 310 315 320

Val Ile Ser Val Val Thr Thr Gly Leu Asp Arg Glu Ser Phe Pro Thr
325 330 335

Tyr Thr Leu Val Val Gln Ala Ala Asp Leu Gln Gly Glu Gly Leu Ser
340 345 350

Thr Thr Ala Thr Ala Val Ile Thr Val Thr Asp Thr Asn Asp Asn Pro
355 360 365

Pro Ile Phe Asn Pro Thr Thr Tyr Lys Gly Gln Val Pro Glu Asn Glu
370 375 380

Ala Asn Val Val Ile Thr Thr Leu Lys Val Thr Asp Ala Asp Ala Pro
385 390 395 400

Asn Thr Pro Ala Trp Glu Ala Val Tyr Thr Ile Leu Asn Asp Asp Gly
405 410 415

Gly Gln Phe Val Val Thr Thr Asn Pro Val Asn Asn Asp Gly Ile Leu
420 425 430

Lys Thr Ala Lys Gly Leu Asp Phe Glu Ala Lys Gln Gln Tyr Ile Leu
435 440 445

His Val Ala Val Thr Asn Val Val Pro Phe Glu Val Ser Leu Thr Thr
450 455 460

Ser Thr Ala Thr Val Thr Val Asp Val Leu Asp Val Asn Glu Gly Pro
465 470 475 480

Ile Phe Val Pro Pro Glu Lys Arg Val Glu Val Ser Glu Asp Phe Gly
485 490 495

Val Gly Gln Glu Ile Thr Ser Tyr Thr Ala Gln Glu Pro Asp Thr Phe
500 505 510

Met Glu Gln Lys Ile Thr Tyr Arg Ile Trp Arg Asp Thr Arg Asn Trp
515 520 525

Leu Glu Ile Asn Pro Asp Thr Gly Ala Ile Ser Thr Arg Ala Glu Leu
530 535 540

Asp Arg Glu Asp Phe Glu His Val Lys Asn Ser Thr Tyr Thr Ala Leu
545 550 555 560

Ile Ile Ala Thr Asp Asn Gly Ser Pro Val Ala Thr Gly Thr Gly Thr
565 570 575

Leu Leu Leu Ile Leu Ser Asp Val Asn Asp Asn Ala Pro Ile Pro Glu
580 585 590

Pro Arg Thr Ile Phe Phe Cys Glu Arg Asn Pro Lys Pro Gln Val Ile
595 600 605

Asn Ile His Asp Ala Asp Leu Pro Pro Asn Thr Ser Pro Phe Thr Ala
610 615 620

Glu Leu Thr His Gly Arg Val Pro Asn Trp Thr Ile Gln Tyr Asn Asp
625 630 635 640

Pro Thr Gln Glu Ser Ile Ile Leu Lys Pro Lys Met Ala Leu Glu Val

645 650 655

Gly Asp Tyr Lys Ile Asn Leu Lys Leu Met Asp Asn Gln Asn Lys Asp
660 665 670

Gln Val Thr Thr Leu Glu Val Ser Val Cys Asp Cys Glu Gly Ala Ala
675 680 685

Gly Val Cys Arg Lys Ala Gln Pro Val Glu Ala Gly Leu Gln Ile Pro
690 695 700

Ala Ile Leu Gly Ile Leu Gly Gly Ile Leu Ala Leu Leu Ile Leu Ile
705 710 715 720

Leu Leu Leu Leu Leu Phe Leu Arg Arg Arg Ala Val Val Lys Glu Pro
725 730 735

Leu Leu Pro Pro Glu Asp Asp Thr Arg Asp Asn Val Tyr Tyr Tyr Asp
740 745 750

Glu Glu Gly Gly Gly Glu Glu Asp Gln Asp Phe Asp Leu Ser Gln Leu
755 760 765

His Arg Gly Leu Asp Ala Arg Pro Glu Val Thr Arg Asn Asp Val Ala
770 775 780

Pro Thr Leu Met Ser Val Pro Arg Tyr Leu Pro Arg Pro Ala Asn Pro
785 790 795 800

Asp Glu Ile Gly Asn Phe Ile Asp Glu Asn Leu Lys Ala Ala Asp Thr
805 810 815

Asp Pro Thr Ala Pro Pro Tyr Asp Ser Leu Leu Val Phe Asp Tyr Glu
820 825 830

Gly Ser Gly Ser Glu Ala Ala Ser Leu Ser Ser Leu Asn Ser Ser Glu
835 840 845

Ser Asp Lys Asp Gln Asp Tyr Asp Tyr Leu Asn Glu Trp Gly Asn Pro
850 855 860

Phe Lys Lys Leu Ala Asp Met Tyr Gly Gly Gly Glu Asp His
865 870 875

International Application No: PCT/ /

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 58, lines 5-25 of the description

A. IDENTIFICATION OF DEPOSIT

Further deposits are identified on an additional sheet

Name of depositary institution

American Type Culture Collection

Address of depositary institution (including postal code and country)

12301 Parklawn Drive
Rockville, MD 20852
US

Date of deposit November 12, 1992 Accession Number 69122

B. ADDITIONAL INDICATIONS (leave blank if not applicable). This information is continued on a separate attached sheet

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)

D. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☐ This sheet was received with the International application when filed (to be checked by the receiving Office)

ELNORA Y. RIVERA
INTERNATIONAL DIVISION
(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

International Application No: PCT/ /

Form PCT/RO/134 (cont.)

American Type Culture Collection

12301 Parklawn Drive
Rockville, MD 20852
US

Accession No.	Date of Deposit
69123	November 12, 1992

WHAT IS CLAIMED IS:

1. A purified human E-cadherin protein having the amino acid sequence depicted in Figure 3 (SEQ ID NO:2) from amino acid numbers 1-878.

2. A purified human E-cadherin protein having the amino acid sequence depicted in Figure 3 (SEQ ID NO:2) from amino acid numbers 151-878, which is free of detergents.

3. The protein of claim 1 which is not glycosylated.

4. A purified human E-cadherin protein having the amino acid sequence depicted in Figure 3 (SEQ ID NO:2) from amino acid numbers 151-878, which is not glycosylated.

5. A purified protein comprising a fragment of a human E-cadherin protein consisting of at least 30 sequential amino acids of the human E-cadherin sequence shown in Figure 3 from amino acid numbers 308-878.

6. The protein of claim 5 which displays one or more functional activities associated with a full-length human E-cadherin protein.

7. The protein of claim 6 in which the protein is able to be bound by an antibody to a human E-cadherin protein.

8. A purified protein comprising amino acid numbers 1-150 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

9. A purified protein comprising amino acid numbers 178-289 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

10. A purified protein comprising amino acid numbers 290-401 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

11. A purified protein comprising amino acid numbers 402-513 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

12. A purified protein comprising amino acid numbers 178-513 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

13. A purified protein comprising amino acid numbers 151-703 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

14. A purified protein comprising amino acid numbers 1-703 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

15. A purified protein comprising amino acid numbers 728-878 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

16. A purified protein comprising amino acid numbers 704-878 as depicted in Figure 3 (SEQ ID NO:2), with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

17. A purified protein comprising the E-cadherin amino acid sequence depicted in Figure 3 (SEQ ID NO:2) encoded by nucleotide numbers 1-1053, 510-2686, 1332-3000, 540-1500, 348-906, 890-1648, 384-1208, 641-2046, 685-1336, 880-1661, 1199-1742, 1373-1742, 1705-2204, or 2458-2775, with the proviso that said protein is not a mature human E-cadherin protein comprising amino acids 151-878 of Figure 3.

18. A fragment of a human E-cadherin protein, said fragment consisting of at least ten sequential amino acids selected from the repeat region or cytoplasmic domain of a human E-cadherin protein, said human E-cadherin protein having the amino acid sequence shown in Figure 3 (SEQ ID NO:2), in which said fragment is able to be bound by antibody to said human E-cadherin protein.

19. A chimeric protein comprising a functionally active fragment of a human E-cadherin protein joined via a peptide bond to an amino acid sequence of a protein other than a human E-cadherin protein, in which the fragment of the human E-cadherin protein is selected from the group consisting of the extracellular domain, the cytoplasmic domain, the repeat region, and the conserved cysteine domain.

20. A chimeric protein comprising a functionally active fragment of a human E-cadherin protein joined via a peptide bond to an amino acid sequence of a protein other than a human E-cadherin protein, in which the fragment of the human E-cadherin protein consists of at least 30 sequential amino acids of the human E-cadherin sequence shown in Figure 3 from amino acid numbers 308-878.

21. The protein according to claim 20 in which the at least 30 sequential amino acids are from the cytoplasmic domain of the human E-cadherin protein.

22. A monoclonal antibody which binds to a human E-cadherin protein and which does not bind to a mouse or chicken E-cadherin protein.

23. An antibody which binds to the cytoplasmic domain of a human E-cadherin protein.

24. A purified nucleic acid encoding a human E-cadherin protein having the amino acid sequence shown in Figure 3 (SEQ ID NO:2) from amino acid numbers 1-878.

25. A purified nucleic acid encoding a human E-cadherin protein having the amino acid sequence shown in Figure 3 (SEQ ID NO:2) from amino acid numbers 151-878.

26. The nucleic acid of claim 25 which lacks introns.

27. A purified nucleic acid which encodes the protein of claim 8, 9, or 10.

28. A purified nucleic acid which encodes the protein of claim 11, 13 or 14.

29. A purified nucleic acid which encodes the protein of claim 15 or 16.

30. A purified nucleic acid having the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1) from nucleotide numbers 116-2749.

31. A purified nucleic acid having the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1) from nucleotide numbers 566-2749.

32. A purified nucleic acid comprising the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1) from nucleotide numbers 1-1053, 510-2686, 1332-3000, 540-1500, 348-906, 890-1648, 384-1208, 641-2046, 685-1336, 880-1661, 1199-1742, 1373-1742, 1705-2204, or 2458-2775.

33. A purified nucleic acid comprising the human E-cadherin nucleotide sequence contained in plasmid bsFLEC, as deposited with the ATCC and assigned accession number 69123.

34. A purified nucleic acid comprising the human E-cadherin nucleotide sequence contained in plasmid bsL5.1, as deposited with the ATCC and assigned accession number 69122.

35. A purified cDNA encoding a protein comprising a fragment of a human E-cadherin protein consisting of at least 30 sequential amino acids of the human E-cadherin sequence depicted in Figure 3 (SEQ ID NO:3) selected from amino acid numbers 308-878.

36. A purified nucleic acid comprising a nucleotide sequence 100% complementary to at least 30 sequential nucleotides of the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1), with the proviso that said nucleotides do not consist of a portion of the nucleotide sequence from nucleotides numbers 617-1036 depicted in Figure 3.

37. A nucleic acid encoding the chimeric protein of claim 20.

38. A nucleic acid vector comprising the nucleic acid of claim 24 or 25.

39. A nucleic acid vector comprising the nucleic acid of claim 27.

40. A cell containing the nucleic acid vector of claim 37.

41. A cell containing the nucleic acid vector of claim 38.

42. A cell containing the nucleic acid vector of claim 39.

43. A cell containing a nucleic acid comprising (a) a first nucleotide sequence encoding human E-cadherin or an at least 30 amino acid functional fragment thereof, with the proviso that said fragment does not consist of amino acids numbers 153-307 depicted in Figure 3 (SEQ ID NO:2); and (b) a promoter operatively linked to the first nucleotide sequence, with the proviso that said promoter is not a human E-cadherin gene promoter.

44. The cell of claim 43 which is a tumor cell.

45. A method for producing a human E-cadherin protein comprising growing the recombinant cell of claim 40, such that the human E-cadherin protein is expressed by the cell; and recovering the expressed human E-cadherin protein.

46. A method for producing a human E-cadherin protein comprising growing the recombinant cell of claim 41, such that the human E-cadherin protein is expressed by the cell; and recovering the expressed human E-cadherin protein.

47. A purified protein which is the product of the method of claim 45.

48. A purified protein which is the product of the method of claim 46.

49. A pharmaceutical composition comprising a therapeutically effective amount of the protein of claim 1; and a pharmaceutically acceptable carrier.

50. A pharmaceutical composition comprising a therapeutically effective amount of the protein of claim 2; and a pharmaceutically acceptable carrier.

51. A pharmaceutical composition comprising a therapeutically effective amount of a purified human E-cadherin protein having the amino acid sequence depicted in Figure 3 (SEQ ID NO:2) from amino acid numbers 151-878; and a pharmaceutically acceptable carrier.

52. A pharmaceutical composition comprising a therapeutically effective amount of the protein of claim 4; and a pharmaceutically acceptable carrier.

53. A pharmaceutical composition comprising a therapeutically effective amount of the protein of claim 12; and a pharmaceutically acceptable carrier.

54. A pharmaceutical composition comprising a therapeutically effective amount of the antibody of claim 22; and a pharmaceutically acceptable carrier.

55. A pharmaceutical composition comprising a therapeutically effective amount of the nucleic acid of claim 24, 25, or 31; and a pharmaceutically acceptable carrier.

56. A pharmaceutical composition comprising a therapeutically effective amount of the cDNA of claim 35; and a pharmaceutically acceptable carrier.

57. A method of treating or preventing a malignancy in a subject comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the protein of claim 1.

58. A method of treating or preventing a malignancy in a subject comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the protein of claim 2.

59. A method of treating or preventing a malignancy in a subject comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the protein of claim 4.

60. A method of treating or preventing a malignancy in a subject comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the protein of claim 17.

61. A method of treating or preventing a malignancy in a subject comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the nucleic acid of claim 24 or 25.

62. A method of treating a benign dysproliferative disorder in a subject comprising administering to a subject in need of such treatment a therapeutically effective amount of the protein of claim 1.

63. A method of treating a benign dysproliferative disorder in a subject comprising administering to a subject in need of such treatment a therapeutically effective amount of the protein of claim 2.

64. A method of treating a benign dysproliferative disorder in a subject comprising administering to a subject in need of such treatment a therapeutically effective amount of the protein of claim 5.

65. A method of treating a benign dysproliferative disorder in a subject comprising administering to a subject in need of such treatment a therapeutically effective amount of the nucleic acid of claim 24 or 25.

66. A purified nucleic acid comprising a nucleotide sequence 100% complementary to at least a 30 sequential nucleotide portion of the nucleotide sequence as depicted in Figure 3 (SEQ ID NO:1).

67. The nucleic acid of claim 66 which is complementary to a portion of the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1) selected from nucleotides numbers 116-2748.

68. A method of detecting metastatic potential in a cell comprising detecting or measuring the level of human E-cadherin in the cell, in which a change in localization or decreased level of human E-cadherin relative to the localization or level of human E-cadherin in a non-malignant cell indicates that the cell has metastatic potential; wherein the detection or measurement of human E-cadherin is carried out by a method comprising contacting the cell with an antibody to the cytoplasmic domain of human E-cadherin such that immunospecific binding can occur, and detecting or measuring any immunospecific binding to the antibody

69. A method of treating or preventing a malignancy or a benign dysproliferative disorder in a subject comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the cell of claim 43.

70. A method of promoting nerve or tissue regeneration in a subject comprising administering to a subject in need of such treatment a therapeutically effective amount of the antibody of claim 23.

71. A method of promoting nerve or tissue regeneration in a subject comprising administering to a subject in need of such treatment a therapeutically effective amount of the nucleic acid of claim 66.

72. A fragment of the antibody of claim 23 containing the antibody binding domain.

73. A method of promoting wound healing in a subject comprising delivering to the site of a wound in a subject a therapeutically effective amount of the protein of claim 2.

74. A method of promoting wound healing in a subject comprising delivering to the site of a wound in a subject a therapeutically effective amount of the nucleic acid of claim 24 or 25.

75. A method of treating an inflammatory disorder comprising delivering to the site of inflammation in a subject a therapeutically effective amount of the protein of claim 2.

76. A method of treating an inflammatory disorder comprising delivering to the site of inflammation in a subject a therapeutically effective amount of the nucleic acid of claim 24 or 25.

77. A method of treating or preventing gestational disease or fetal wastage comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the protein of claim 2.

78. A method of treating or preventing gestational disease or fetal wastage comprising administering to a subject in need of such treatment or prevention a therapeutically or prophylactically effective amount of the nucleic acid of claim 24 or 25.

79. A kit comprising in one or more containers:

(a) a first oligonucleotide of at least 15 nucleotides in size, consisting of a nucleotide sequence that is a first portion of the nucleotide sequence depicted in Figure 3 (SEQ ID NO:1), wherein said first portion is not identical to or contained within (i) nucleotide numbers 617-1036 depicted in Figure 3 (SEQ ID NO:1), or (ii) the mouse or chicken E-cadherin coding sequence; and

(b) a second oligonucleotide of at least 15 nucleotides in size, consisting of a nucleotide sequence that (i) is complementary to a second portion of the nucleotide sequence depicted in Figure 3, said second portion situated 3′ to said first portion.

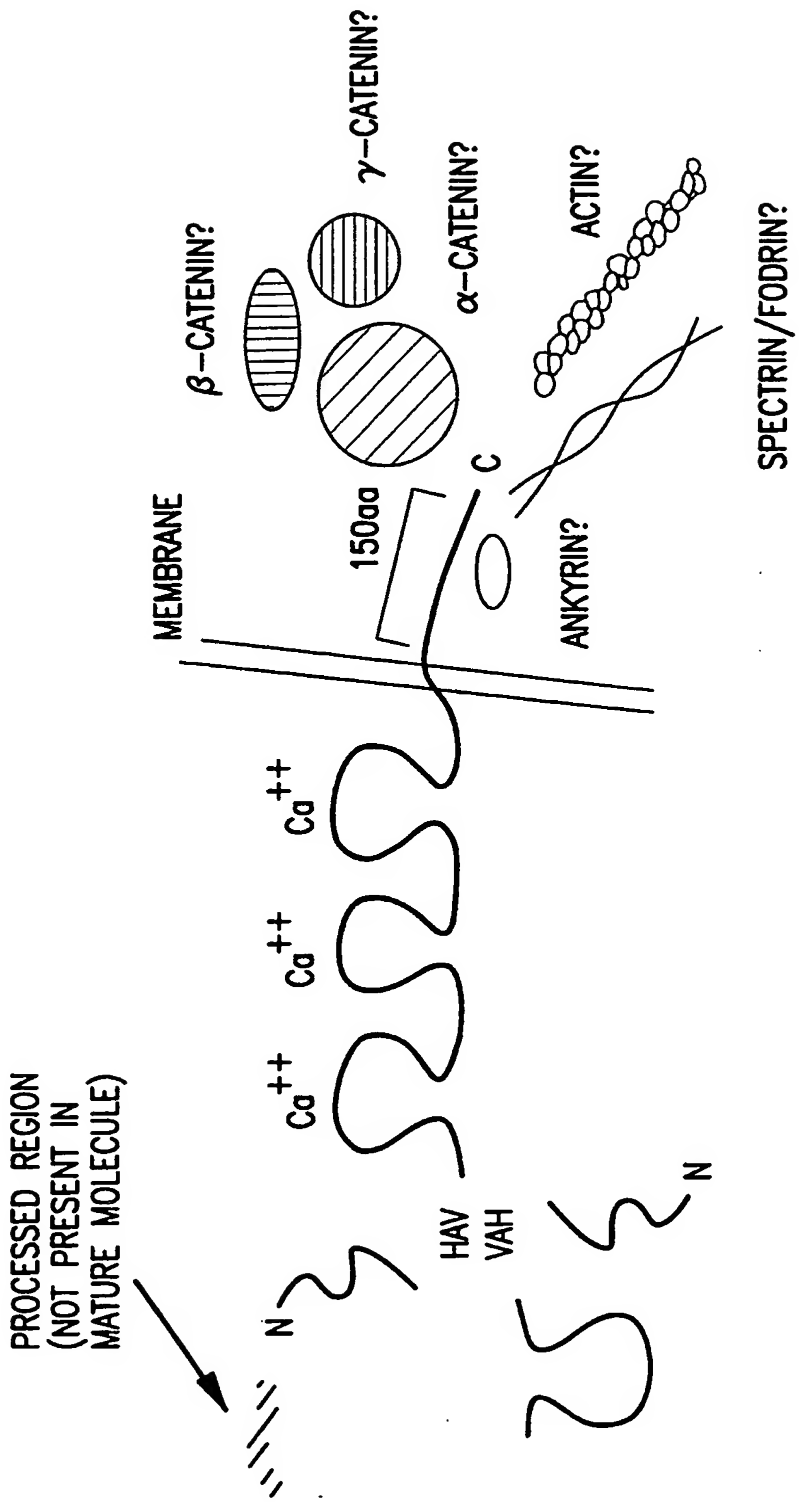
PROCESSED REGION
(NOT PRESENT IN
MATURE MOLECULE)
N
HAV
VAH
N
Ca++
Ca++
Ca++
MEMBRANE
150aa
C
ANKYRIN?
β-CATENIN?
γ-CATENIN?
α-CATENIN?
ACTIN?
SPECTRIN/FODRIN?

FIG.1

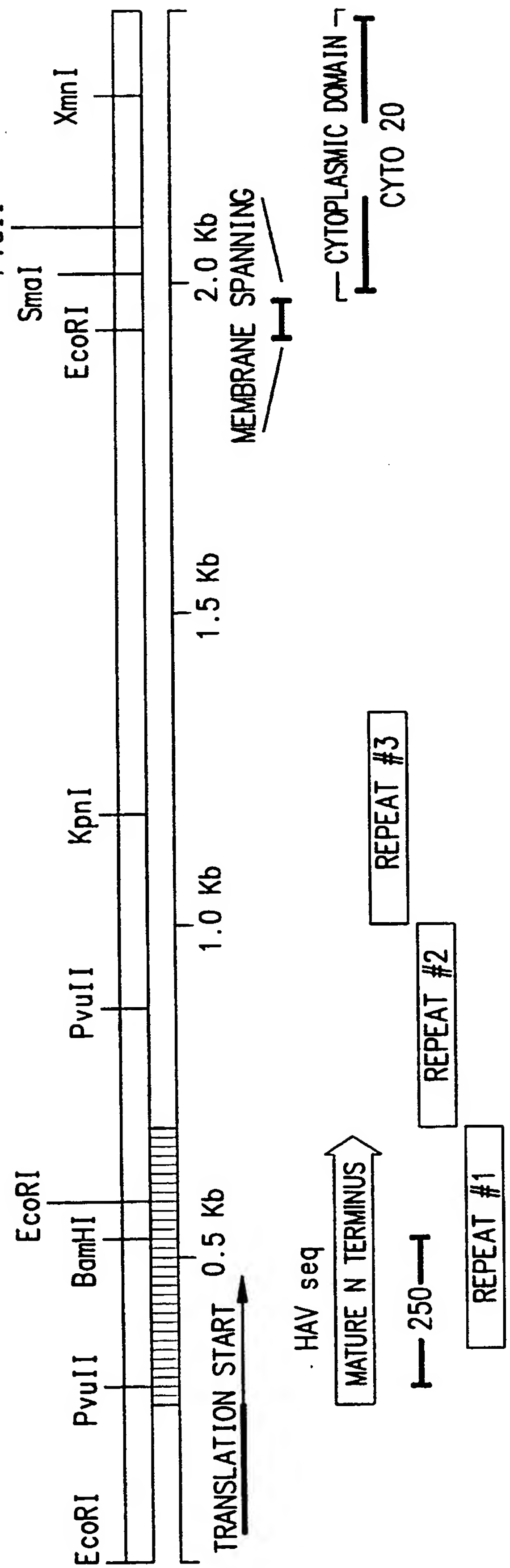
EcoRI
PvuII
BamHI
EcoRI
PvuII
KpnI
EcoRI
SmaI
PvuII
XmnI
0.5 Kb
1.0 Kb
1.5 Kb
2.0 Kb
TRANSLATION START
HAV seq
MATURE N TERMINUS
250
REPEAT #1
REPEAT #2
REPEAT #3
MEMBRANE SPANNING
CYTOPLASMIC DOMAIN
CYTO 20

FIG.2

```
GAATTCCGGAAAGCACCTGTGAGCTTGGCAAGTCAGTTCAGAGCTCCAGCCCCGCTCCAGCCCCGGCCCGACCCGACCGCACCCGGCGCCTGCCTCGCTCGGGCTCCCCGGCCAGCCATGGG 120
                                                                                                                   M  G   2
CCCTTGGAGCCCGCAGCCTCTCGGGCCTGCTGCTGCTGCTGAGGTCTCCTCTTGGCTCTCAGGAGCGGAGCCCTCCTCCCTGTTTGACGCGAGAGCTACACGTTCACGGTGCCCCGGCGCC 240
 P  W  S  R  S  L  S  G  L  L  L  L  L  R  S  P  L  G  S  Q  E  R  S  P  P  P  C  L  T  R  E  L  H  V  H  G  A  P  A  P  42
ACCTGAGAAGAGGCCCCGCGTCTGGGCAGAGTGAATTTTGAAGATTGCACCGGTCGACAAAGGACAGCTATTTTCCTGACACCGATTCCGAAAGTGGGCACAGATGGTGTGATTACAGTCAA 360
 P  E  K  R  P  R  L  G  R  V  N  F  E  D  C  T  G  R  Q  R  T  A  I  F  L  T  P  I  P  K  V  G  T  D  G  V  I  T  V  K  82
AAGGCCTCTACGGTTTCATAACCCAACAGATCCATTTCTTGGTCTACGCTGGGACTCCACCTACAGAAAGTTTTCCACCAAAGTCACGCTGAATACAGTGGGGCACCACCACCGCCCCCC 480
 R  P  L  R  F  H  N  P  T  D  P  F  L  G  L  R  W  D  S  T  Y  R  K  F  S  T  K  V  T  L  N  T  V  G  H  H  H  R  P  P  122
GCCCCATCAGGCCTCCGTTTCTGGAATCCAAGCAGAATTGCTCACATTTCCCAACTCCTCTCCTGGCCTCAGAAGACAGAAGAGAGACTGGGTTATTCCTCCCATCAGCTGCCCAGAAAA 600
 P  H  Q  A  S  V  S  G  I  Q  A  E  L  L  T  F  P  N  S  S  P  G  L  R  R  Q  K  R  D  W  V  I  P  P  I  S  C  P  E  N  162
TGAAAAAGGCCCATTTCCTAAAAACCTGGTTCAGATCAAATCCAACAAAGACAAAGAAGGCAAGGTTTTCTACAGCATCACTGGCCAAGGAGCTGACACACCCCCTGTTGGTGTCTTTAT 720
 E  K  G  P  F  P  K  N  L  V  Q  I  K  S  N  K  D  K  E  G  K  V  F  Y  S  I  T  G  Q  G  A  D  T  P  P  V  G  V  F  I  202
TATTGAAAGAGAAACAGGATGGCTGAAGGTGACAGAGCCTCTGGATAGAGAACGCATTGCCACATACACTCTCTTCTCTCACGCTGTGTCATCCAACGGGAATGCAGTTGAGGATCCAAT 840
 I  E  R  E  T  G  W  L  K  V  T  E  P  L  D  R  E  R  I  A  T  Y  T  L  F  S  H  A  V  S  S  N  G  N  A  V  E  D  P  M  242
```

FIG.3A

GGAGATTTTGATCACGGTAACCGATCAGAATGACAACAAGCCCGAATTCACCCAGGAGGTCTTTAAGGGGTCTGTCATGGAAGGTGCTCTTCCAGGAACCTCTGTGATGGAGGTCACAGC 960

E I L I T V T D Q N D N K P E F T Q E V F K G S V M E G A L P G T S V M E V T A 282

CACAGACGCGGACGATGATGTGAACACCTACAATGCCGCCATCGCTTACACCATCCTCAGCCAAGATCCTGAGCTCCCTGACAAAAATATGTTCACCATTAACAGGAACACAGGAGTCAT 1080

T D A D D D V N T Y N A A I A Y T I L S Q D P E L P D K N M F T I N R N T G V I 322

CAGTGTGGTCACCACTGGGCTGGACCGAGAGAGTTTCCCTACGTATACCCTGGTGGTTCAAGCTGCTGACCTTCAAGGTGAGGGGTTAAGCACAACAGCAACAGCTGTGATCACAGTCAC 1200

S V V T T G L D R E S F P T Y T L V V Q A A D L Q G E G L S T T A T A V I T V T 362

TGACACCAACGATAATCCTCCGATCTTCAATCCCACCACGTACAAGGGTCAGGTGCCTGAGAACGAGGCTAACGTCGTAATCACCACACTGAAAGTGACTGATGCTGATGCCCCCAATAC 1320

D T N D N P P I F N P T T Y K G Q V P E N E A N V V I T T L K V T D A D A P N T 402

CCCAGCGTGGGAGGCTGTATACACCATATTGAATGATGATGGTGGACAATTTGTCGTCACCACAAATCCAGTGAACAACGATGGCATTTTGAAAACAGCAAAGGGCTTGGATTTTGAGGC 1440

P A W E A V Y T I L N D D G G Q F V V T T N P V N N D G I L K T A K G L D F E A 442

CAAGCAGCAGTACATTCTACACGTAGCAGTGACGAATGTGGTACCTTTTGAGGTCTCTCTCACCACCTCCACAGCCACCGTCACCGTGGATGTGCTGGATGTGAATGAAGGCCCCATCTT 1560

K Q Q Y I L H V A V T N V V P F E V S L T T S T A T V T V D V L D V N E G P I F 482

TGTGCCTCCTGAAAAGAGAGTGGAAGTGTCCGAGGACTTTGGCGTGGGCCAGGAAATCACATCCTACACTGCCCAGGAGCCAGACACATTTATGGAACAGAAAATAACATATCGGATTTG 1680

V P P E K R V E V S E D F G V G Q E I T S Y T A Q E P D T F M E Q K I T Y R I W 522

FIG.3B

```
GAGAGACACTCGCAACTGGCTGGAGATTAATCCGGACACTGGTGCCATTTCCACTCGGGCTGAGCTGGACAGGGAGGATTTTGAGCACGTGAAGAACAGCACGTACACAGCCCTAATCAT  1800
 R  D  T  R  N  W  L  E  I  N  P  D  T  G  A  I  S  T  R  A  E  L  D  R  E  D  F  E  H  V  K  N  S  T  Y  T  A  L  I  I   562

AGCTACAGACAATGGTTCTCCAGTTGCTACTGGAACAGGGACACTTCTGCTGATCCTGTCTGATGTGAATGACAACGCCCCCATACCAGAACCTCGAACTATATTCTTCTGTGAGAGGAA  1920
 A  T  D  N  G  S  P  V  A  T  G  T  G  T  L  L  L  I  L  S  D  V  N  D  N  A  P  I  P  E  P  R  T  I  F  F  C  E  R  N   602

TCCAAAGCCTCAGGTCATAAACATTCATGATGCAGACCTTCCTCCCAATACATCTCCCTTCACAGCAGAACTAACACACGGGCGAGTGCCCAACTGGACCATTCAGTACAACGACCCAAC  2040
 P  K  P  Q  V  I  N  I  H  D  A  D  L  P  P  N  T  S  P  F  T  A  E  L  T  H  G  R  V  P  N  W  T  I  Q  Y  N  D  P  T   642

CCAAGAATCTATCATTTTGAAGCCAAAGATGGCCTTAGAGGTGGGTGACTACAAAATCAATCTCAAGCTCATGGATAACCAGAATAAAGACCAAGTGACCACCTTAGAGGTCAGCGTGTG  2160
 Q  E  S  I  I  L  K  P  K  M  A  L  E  V  G  D  Y  K  I  N  L  K  L  M  D  N  Q  N  K  D  Q  V  T  T  L  E  V  S  V  C   682

TGACTGTGAAGGGGCCGCCGGCGTCTGTAGGAAGGCACAGCCTGTCGAAGCAGGATTGCAAATTCCTGCCATTCTGGGGATTCTTGGAGGAATTCTTGCTTTGCTAATTCTGATTCTGCT  2280
 D  C  E  G  A  A  G  V  C  R  K  A  Q  P  V  E  A  G  L  Q  I  P  A  I  L  G  I  L  G  G  I  L  A  L  L  I  L  I  L  L   722

GCTCTTGCTGTTTCTTCGGAGGAGAGCGGTGGTCAAAGAGCCCTTACTGCCCCCAGAGGATGACACCCGGGACAACGTTTATTACTATGATGAAGAAGGAGGCGGAGAAGAGGACCAGGA  2400
 L  L  L  F  L  R  R  R  A  V  V  K  E  P  L  L  P  P  E  D  D  T  R  D  N  V  Y  Y  Y  D  E  E  G  G  G  E  E  D  Q  D   762
```

FIG.3C

```
CTTTGACTTGAGCCAGCTGCACAGGGGCCTGGACGCTCGGCCTGAAGTGACTCGTAACGACGTTGCACCAACCCTCATGAGTGTCCCCCGGTATCTTCCCCGCCCTGCCAATCCCGATGA  2520
 F  D  L  S  Q  L  H  R  G  L  D  A  R  P  E  V  T  R  N  D  V  A  P  T  L  M  S  V  P  R  Y  L  P  R  P  A  N  P  D  E   802

AATTGGAAATTTTATTGATGAAAATCTGAAAGCGGCTGATACTGACCGCACAGCCCCGCCTTATGATTCTCTGCTCGTGTTTGACTATGAAGGAAGCGGTTCCGAAGCTGCTAGTCTGAG  2640
 I  G  N  F  I  D  E  N  L  K  A  A  D  T  D  P  T  A  P  P  Y  D  S  L  L  V  F  D  Y  E  G  S  G  S  E  A  A  S  L  S   842

CTCCCTGAACTCCTCAGAGTCAGACAAAGACCAGGACTATGACTACTTGAACGAATGGGGCAATCCGTTCAAGAAGCTGGCTGACATGTACGGAGGCGGCGAGGACCACTAGGGGACTCG  2760
 S  L  N  S  S  E  S  D  K  D  Q  D  Y  D  Y  L  N  E  W  G  N  P  F  K  K  L  A  D  M  Y  G  G  G  E  D  H

AGAGAGGCGG CCCAGACCAT GTGCAGAAAT GCAGAATCA GCGTTCTGTT GTTTT                                                            2815
```

FIG.3D

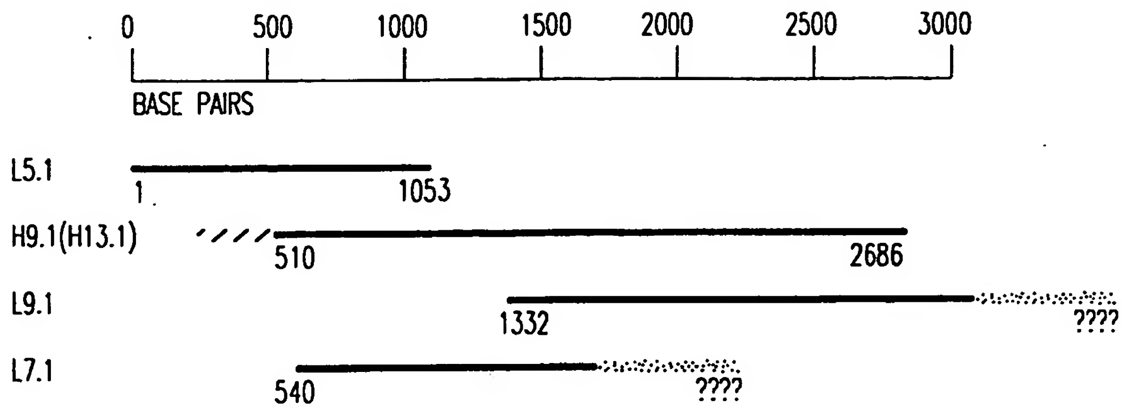
0
500
1000
1500
2000
2500
3000
BASE PAIRS
L5.1
1
1053
H9.1(H13.1)
510
2686
L9.1
1332
????
L7.1
540
????

FIG.4A

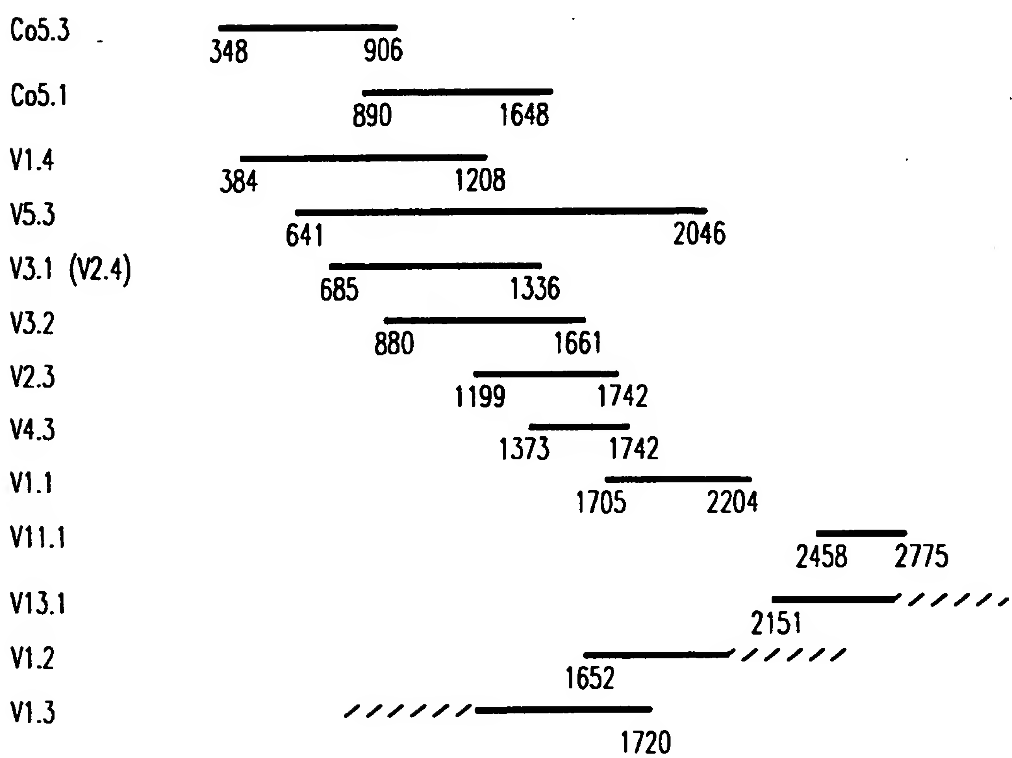
Co5.3
348
906
Co5.1
890
1648
V1.4
384
1208
V5.3
641
2046
V3.1 (V2.4)
685
1336
V3.2
880
1661
V2.3
1199
1742
V4.3
1373
1742
V1.1
1705
2204
V11.1
2458
2775
V13.1
2151
V1.2
1652
V1.3
1720

FIG.4B

EcoR I Sac I

GAATTCCGGAAAGCACCTGTGAGCTTGGCAAGTCAGTTCAGAGCTCCAGCCCGCTCCAGCCCGGCCCGACCCGACCGCAC 80

Nar I Nco I

CCGGCGCCTGCCTCGCTCGGGCTCCCCGGCCAGCCATGGGCCCTTGGAGCCCGCAGCCTCTCGGGCCTGCTGCTGCTGCTG 160

Translation Start

M G P W S R S L S G L L L L L

Nar I

AGGTCTCCTCTTGGCTCTCAGGAGCGGAGCCCTCCTCCCTGTTTGACGCGAGAGCTACACGTTCACGGTGCCCCGGCGCC 240

R S P L G S Q E R S P P P C L T R E L H V H G A P A P

Sal I
Hinc II

ACCTGAGAAGAGGCCGCGTCTGGGCAGAGTGAATTTTGAAGATTGCACCGGTCGACAAAGGACAGCTATTTTCCTGACAC 320

P E K R P R L G R V N F E D C T G R Q R T A I F L T

Stu I

CGATTCCGAAAGTGGGCACAGATGGTGTGATTACAGTCAAAAGGCCTCTACGGTTTCATAACCCAACAGATCCATTTCTT 400

P I P K V G T D G V I T V K R P L R F H N P T D P F L

Xmn I

GGTCTACGCTGGGACTCCACCTACAGAAAGTTTTCCACCAAAGTCACGCTGAATACAGTGGGGCACCACCACCGCCCCCC 480

G L R W D S T Y R K F S T K V T L N T V G H H H R P P

FIG.5A

Stu I

GCCCCATCAGGCCTCCGTTTCTGGAATCCAAGCAGAATTGCTCACATTTCCCAACTCCTCTCCTGGCCTCAGAAGACAGA 560

P H Q A S V S G I Q A E L L T F P N S S P G L R R Q

Pvu II

AGAGAGACTGGGTTATTCCTCCCATCAGCTGCCCAGAAAATGAAAAAGGCCCATTTCCTAAAAACCTGGTTCAGATCAAA 640

Mature N Terminus

K R D W V I P P I S C P E N E K G P F P K N L V Q I K

TCCAACAAAGACAAAGAAGGCAAGGTTTTCTACAGCATCACTGGCCAAGGAGCTGACACACCCCCTGTTGGTGTCTTTAT 720

S N K D K E G K V F Y S I T G G G A D T P P V G V F I

TATTGAAAGAGAAACAGGATGGCTGAAGGTGACAGAGCCTCTGGATAGAGAACGCATTGCCACATACACTCTCTTCTCTC 800

Rec

I E R E T G W L K V T E P L D R E R I A T Y T L F S

Bam HI BstE II

ACGCTGTGTCATCCAACGGGAATGCAGTTGAGGATCCAATGGAGATTTTGATCACGGTAACCGATCAGAATGACAACAAG 880

Rec Seq

H A V S S N G N A V E D P M E I L I T V T D Q N D N K

EcoRI

CCCGAATTCACCCAGGAGGTCTTTAAGGGGTCTGTCATGGAAGGTGCTCTTCCAGGAACCTCTGTGATGGAGGTCACAGC 960

P E F T Q E V F K G S V M E G A L P G T S V M E V T A

FIG.5B

Sac I

CACAGACGCCGGACGATGATGTGAACACCTACAATGCCGCCATCGCTTACACCATCCTCAGCCAAGATCCTGAGCTCCCTG 1040

T D A D D D V N T Y N A A I A Y T I L S Q D P E L P

BstE II BstX I

ACAAAAATATGTTCACCATTAACAGGAACACAGGAGTCATCAGTGTGGTCACCACTGGGCTGGACCGAGAGAGTTTCCCT 1120

D K N M F T I N R N T G V I S V V T T G L D R E S F P

Pvu II

ACGTATACCCTGGTGGTTCAAGCTGCTGACCTTCAAGGTGAGGGGTTAAGCACAACAGCAACAGCTGTGATCACAGTCAC 1200

T Y T L V V Q A A D L Q G E G L S T T A T A V I T V T

TGACACCAACGATAATCCTCCGATCTTCAATCCCACCACGTACAAGGGTCAGGTGCCTGAGAACGAGGCTAACGTCGTAA 1280

D T N D N P P I F N P T T Y K G Q V P E N E A N V V

TCACCACACTGAAAGTGACTGATGCTGATGCCCCCAATACCCCAGCGTGGGAGGCTGTATACACCATATTGAATGATGAT 1360

I T T L K V T D A D A P N T P A W E A V Y T I L N D D

GGTGGACAATTTGTCGTCACCACAAATCCAGTGAACAACGATGGCATTTTGAAAACAGCAAAGGGCTTGGATTTTGAGGC 1440

G G Q F V V T T N P V N N D G I L K T A K G L D F E A

Kpn I

CAAGCAGCAGTACATTCTACACGTAGCAGTGACGAATGTGGTACCTTTTGAGGTCTCTCTCACCACCTCCACAGCCACCG 1520

K Q Q Y I L H V A V T N V V P F E V S L T T S T A T

TCACCGTGGATGTGCTGGATGTGAATGAAGGCCCCATCTTTGTGCCTCCTGAAAAGAGAGTGGAAGTGTCCGAGGACTTT 1600

V T V D V L D V N E G P I F V P P E K R V E V S E D F

GGCGTGGGCCAGGAAATCACATCCTACACTGCCCAGGAGCCAGACACATTTATGGAACAGAAAATAACATATCGGATTTG 1680

G V G Q E I T S Y T A Q E P D T F M E Q K I T Y R I W

FIG.5C

GAGAGACACTCGCAACTGGCTGGAGATTAATCCGGACACTGGTGCCATTTCCACTCGGGCTGAGCTGGACAGGGAGGATT 1760

R D T R N W L E I N P D T G A I S T R A E L D R E D

TTGAGCACGTGAAGAACAGCACGTACACAGCCCTAATCATAGCTACAGACAATGGTTCTCCAGTTGCTACTGGAACAGGG 1840

F E H V K N S T Y T A L I I A T D N G S P V A T G T G

ACACTTCTGCTGATCCTGTCTGATGTGAATGACAACGCCCCCATACCAGAACCTCGAACTATATTCTTCTGTGAGAGGAA 1920

T L L L I L S D V N D N A P I P E P R T I F F C E R N

TCCAAAGCCTCAGGTCATAAACATTCATGATGCAGACCTTCCTCCCAATACATCTCCCTTCACAGCAGAACTAACACACG 2000

P K P Q V I N I H D A D L P P N T S P F T A E L T H

GGCGAGTGCCCAACTGGACCATTCAGTACAACGACCCAACCCAAGAATCTATCATTTTGAAGCCAAAGATGGCCTTAGAG 2080

G R V P N W T I Q Y N D P T Q E S I I L K P K M A L E

GTGGGTGACTACAAAATCAATCTCAAGCTCATGGATAACCAGAATAAAGACCAAGTGACCACCTTAGAGGTCAGCGTGTG 2160

V G D Y K I N L K L M D N Q N K D Q V T T L E V S V C

TGACTGTGAAGGGGCCGCCGGCGTCTGTAGGAAGGCACAGCCTGTCGAAGCAGGATTGCAAATTCCTGCCATTCTGGGGA 2240

Transmembrane seq

D C E G A A G V C R K A Q P V E A G L Q I P A I L G

EcoRI

TTCTTGGAGGAATTCTTGCTTTGCTAATTCTGATTCTGCTGCTCTTGCTGTTTCTTCGGAGGAGAGCGGTGGTCAAAGAG 2320

Transmembrane seq

I L G G I L A L L I L I L L L L L F L R R R A V V K E

FIG.5D

Xma I

CCCTTACTGCCCCCAGAGGATGACACCCGGGACAACGTTTATTACTATGATGAAGAAGGAGGCGGAGAAGAGGACCAGGA 2400

P L L P P E D D T R D N V Y Y Y D E E G G G E E D Q D

Pvu II

CTTTGACTTGAGCCAGCTGCACAGGGGCCTGGACGCTCGGCCTGAAGTGACTCGTAACGACGTTGCACCAACCCTCATGA 2480

F D L S Q L H R G L D A R P E V T R N D V A P T L M

GTGTCCCCCGGTATCTTCCCCGCCCTGCCAATCCCGATGAAATTGGAAATTTTATTGATGAAAATCTGAAAGCGGCTGAT 2560

S V P R Y L P R P A N P D E I G N F I D E N L K A A D

Xmn I

ACTGACCCCACAGCCCCGCCTTATGATTCTCTGCTCGTGTTTGACTATGAAGGAAGCGGTTCCGAAGCTGCTAGTCTGAG 2640

T D P T A P P Y D S L L V F D Y E G S G S E A A S L S

Sac I

CTCCCTGAACTCCTCAGAGTCAGACAAAGACCAGGACTATGACTACTTGAACGAATGGGGCAATCCGTTCAAGAAGCTGG 2720

S L N S S E S D K D Q D Y D Y L N E W G N P F K K L

CTGACATGTACGGAGGCGGCGAGGACCACTAGGGGA 2756

A D M Y G G G E D H G

FIG.5E

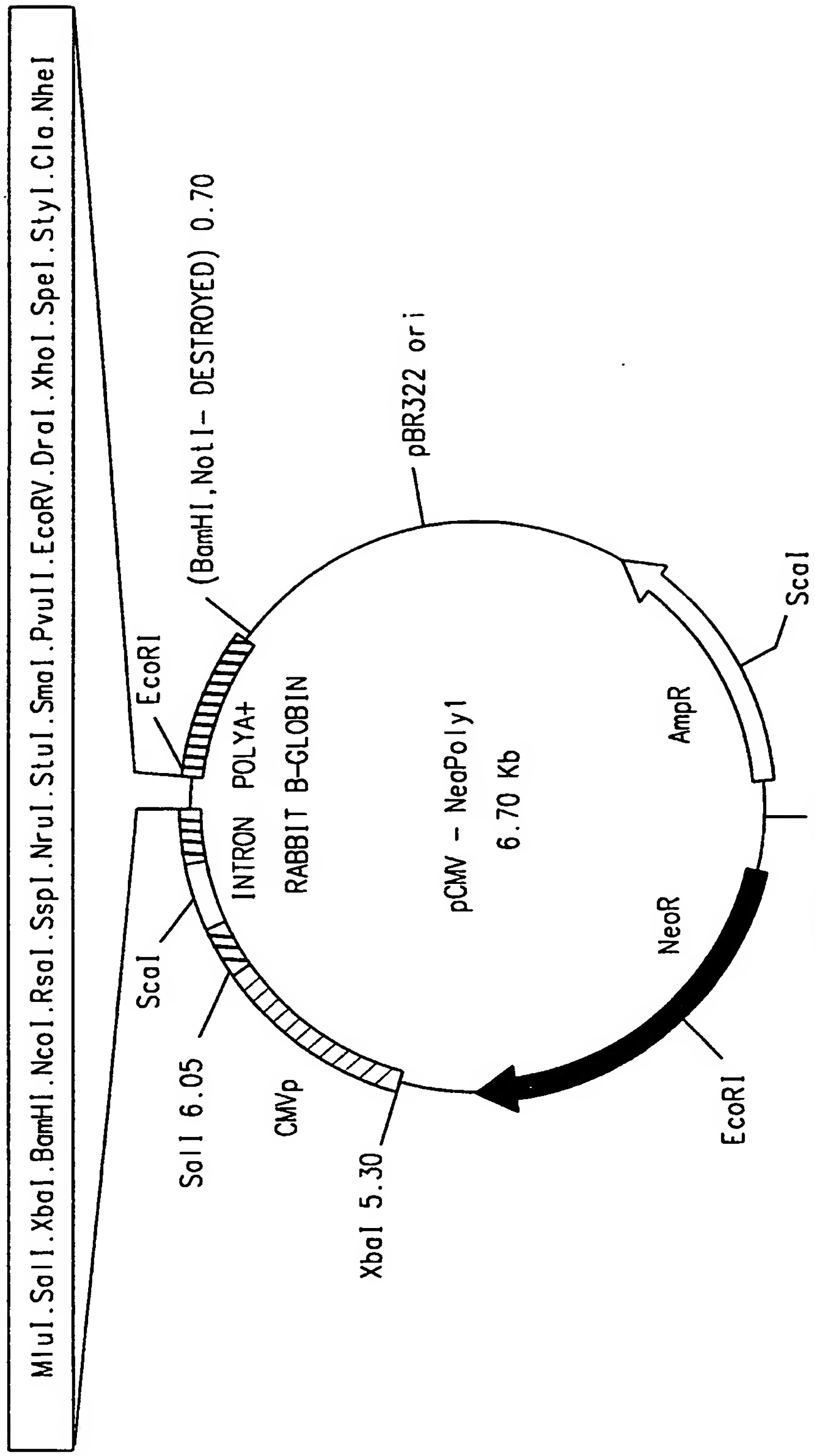
MluI.SalI.XbaI.BamHI.NcoI.RsaI.SspI.NruI.StuI.SmaI.PvuII.EcoRV.DraI.XhoI.SpeI.StyI.ClaI.NheI
EcoRI
(BamHI,NotI- DESTROYED) 0.70
pBR322 ori
ScaI
AmpR
POLYA+
RABBIT B-GLOBIN
INTRON
pCMV - NeoPoly1
6.70 Kb
HindIII 3.40
NeoR
EcoRI
ScaI
SalI 6.05
CMVp
XbaI 5.30

FIG.6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/11097

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350, 3897.1; 514/12, 44; 435/7.1, 69.1, 240.2; 536/23.4, 23.5, 424/85.8, 93R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Differentiation, Vol. 38, No. 1, issued June 1988, A. Mansouri et al., "Characterization and chromosomal location of the gene encoding the human cell adhesion molecule uvomorulin", pages 67-71, see entire document.	1-79
Y	The Journal of Cell Biology, Vol. 108, No. 6, issued June 1989, J. Behrens et al., "Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion, pages 2435-2447, see entire document.	22,23,49-65,68-79

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 February 1994	08 MAR 1994

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Jill Warden for JACQUELINE G. KRIKORIAN
Facsimile No. NOT APPLICABLE	Telephone No. (703) 308-0196

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	The Journal of Cell Biology, Vol. 113, No. 1, issued April 1991, U.H. Frixen et al., "E-cadherin-mediated cell-cell adhesion prevents adhesiveness of human carcinoma cells" pages 173-185, see entire document.	1-23,38-44,49-65, 68-79
Y	Nature, Vol. 329, issued 24 September 1987, A. Nagafuchi et al., "Transformation of cell adhesion properties by exogenously introduced E-cadherin cDNA", pages 341-343, see entire document.	1-79
Y	Cell, Vol. 63, issued 30 November, 1990, M. Ozawa et al., "Single amino acid substitutions in one Ca2+ binding site of uvomorulin abolish the adhesive function", pages 1033-1038, see entire document.	1-79
Y	The EMBO Journal, Vol. 6, No. 12, issued 1987, M. Ringwald et al., "The structure of cell adhesion molecule uvomorulin. Insights into the molecular mechanism of Ca2+-dependent cell adhesion", pages 3647-3653, see entire document.	1-79
Y	Nucleic Acids Research, Vol. 19, No. 23, issued 1991, M. Ringwald et al., "The structure of the gene encoding for the mouse cell adhesion molecule uvomorulin", pages 6533-6539, see entire document.	1-67,72
Y	Development, Vol. 102, issued April 1988, M. Takeichi, "The cadherins: cell-cell adhesion molecules controlling animal morphogenesis", pages 639-655, see entire document.	1-79
Y	WO, A, 92/17608 (FRIXEN et al.), 15 October 1992, see entire document.	1-70,72
Y	WO, A, 91/04745 (RUBIN et al.), 18 April 1991, see entire document.	1-79
Y	Science, Vol. 251, issued 22 March 1991, M. Takeichi, "Cadherin cell adhesion receptors as a morphogenetic regulator", pages 1451-1455, see entire document.	1-79
<u>X</u> Y	Cell, Vol. 34, issued September 1983, C.H. Damsky et al., "Identification and purification of a cell surface glycoprotein mediating intercellular adhesion in embryonic and adult tissue", pages 455-466, see entire document.	<u>1,2,5-7,9-13,18</u> 3,4,8,14-17, 19,22,23,49-54,57-60,62-64,68,69,72

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/11097

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Journal of the American Chemical Society, Vol. 85, issued 1963, R.B. Merrifield, "Solid phase peptide synthesis. I. The synthesis of a tetrapeptide", pages 2149-2154, see entire document.	1-23, 49-53
Y	E. Harlow et al., "Antibodies A Laboratory Manual", published 1988 by Cold Spring Harbor Laboratory (Cold Spring Harbor, New York), pages 72-77, 92-97, 128-135, and 141-157, see entire document.	22,23,54,72
Y	L.S. Goodman et al., "The Pharmacological Basis of Therapeutics", published 1975 by Macmillan Publishing Co., Inc. (New York) pages 1-46, see entire document.	49-67, 69-71,73-78
Y	Cell, Vol. 61, issued 06 April 1990, A. Nose et al., "Localization of specificity determining sites in cadherin cell adhesion molecules", pages 147-155, see entire document.	9-23
Y	Developmental Biology, Vol. 152, issued 1992, K. Shimamura et al. "E-cadherin expression in a particular subset of sensory neurons", pages 242-254, see entire document.	49-56,70,71
X Y	Journal of Cellular Biochemistry, Vol. 34, issued 1987, M.J. Wheelock et al., "Soluble 80-kd fragment of cell-CAM 120/80 disrupts cell-cell adhesion", pages 187-202, see entire document.	2,5-7,9-13,18 3,4,8,14-17, 19,22,23,49-54,57-60,62-64,68,69,72

INTERNATIONAL SEARCH REPORT

International application No. PCT/US93/11097

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

(Telephone Practice)

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (5):

C07K 13/00, 15/28; C07H 21/04; C12N 5/10; C12P 21/00; A61K 37/02, 39/395, 48/00; G01N 33/53

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

530/350, 3897.1; 514/12, 44; 435/7.1, 69.1, 240.2; 536/23.4, 23.5, 424/85.8, 93R

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

Genbank, Medline, Biosis, Embase, CAS, APS, WPI
search terms: cadherins, cell adhesion, recombinant, sequence, chimeric, antibodies, therapy, regeneration, inflammation, wound healing, development

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

I. Claims 1-21, 45-53, 57-60 and 62-64, drawn to a first product, human E-cadherin protein, peptides and chimeric protein, to a method of making said product, and to a first method of use, a method of treating or preventing a malignancy or benign proliferative disorder, classified in Class/subclass 530/350, 435/69.1, and 514/12.

II. Claims 22, 23, 54, 68, and 72, drawn to a second product, an antibody to human E-cadherin, and to a second method of use, a method of detecting metastatic potential in a cell, classified in Class/subclass 530/387.1 and 435/7.1.

III. Claims 24-44, 55, 56, 61, 65, 66, 67, and 79, drawn to a third product, a purified nucleic acid encoding human E-cadherin and chimeric E-cadherin, and cells expressing E-cadherin, and to a third method of use, a method of treating or preventing a malignancy or benign proliferative disorder, classified in Class/subclass 536/23.4, 536/23.5, 435/240.2, and 514/44.

IV. Claim 69, drawn to a method of treating a malignancy using a cell that expresses E-cadherin, classified in Class/subclass 424/93R.

V. Claim 70, drawn to a method of promoting nerve regeneration using an antibody to human E-cadherin, classified in Class/subclass 424/85.8.

VI. Claim 71, drawn to a method of promoting nerve regeneration using the nucleic acid encoding E-cadherin, classified in Class/subclass 514/44.

VII. Claims 73 and 75, drawn to a method of promoting wound healing or treatment of an inflammatory disorder using E-cadherin, classified in Class/subclass 514/12.

VIII. Claims 74 and 76, drawn to a method of promoting wound healing or treatment of an inflammatory disorder using the nucleic acid encoding E-cadherin, classified in Class/subclass 514/44.

IX. Claim 77, drawn to a method of treating of preventing gestational disease or fetal wastage using E-cadherin, classified in Class/subclass 514/12.

X. Claim 78, drawn to a method of treating of preventing gestational disease or fetal wastage using the nucleic acid encoding E-cadherin, classified in Class/subclass 514/44.

The claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept. For example, the products of Groups I-III and the methods of Groups I-X do not have a special technical relationship since the reagents, steps, and objectives are different, and the etiologies and pathologies of the diseases or conditions are different. Note that PCT Rule 13 does not provide for multiple methods within a single application.